U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
2017

U.PORTO

Candidate Genes Related with Development Time Variation in *Drosophila americana* and *D. novamexicana*

José Pedro Taveira Gomes da Silva Barbosa

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Candidate Genes Related with Development Time Variation in *Drosophila americana* and *D. novamexicana*

José Pedro Taveira Gomes da Silva Barbosa

Cell and Molecular Biology Master's Degree Dissertation
Presented to Faculty of Science, University of Porto

2017

MSc

Candidate Genes Related with Development Time Variation in *Drosophila americana* and *D. novamexicana*

José Pedro Taveira Gomes da Silva Barbosa
Cell and Molecular Biology Master's Degree
Science Department, FCUP
2017

Advisor
Jorge Manuel de Sousa Basto Vieira
Main researcher at IBMC
jbvieira@ibmc.up.pt

Coadvisor
Cristina Alexandra Gonçalves Paula Vieira
Auxiliary researcher at IBMC
cgvieira@ibmc.up.pt

Institute Address
IBMC – Instituto Biologia Molecular e Celular
Edifício i3s - Rua Alfredo Allen nº 208
4200-135 Porto, Portugal
220408800

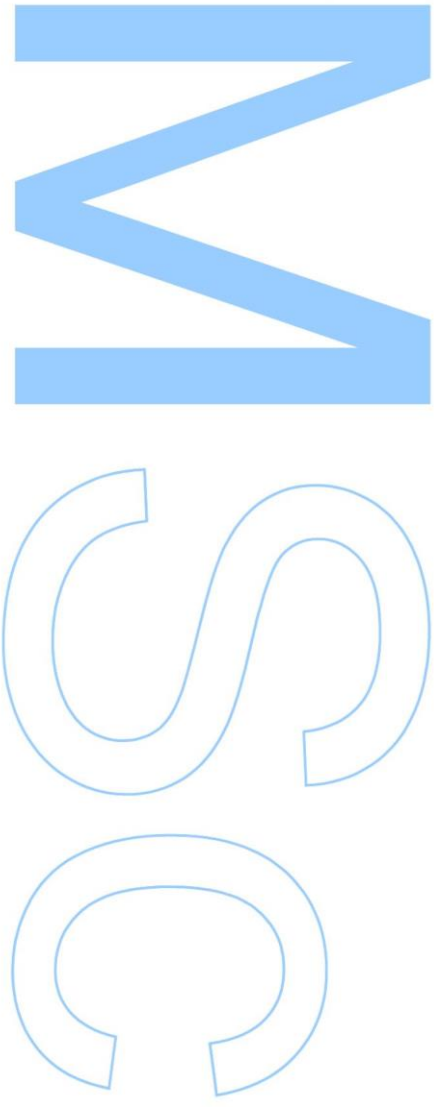
U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Declaração

Eu, José Pedro Taveira Gomes da Silva Barbosa, aluno com o número 201502150, inscrito no mestrado de Biologia Celular e Molecular no presente ano lectivo de 2016/2017, declaro por minha honra que sou o autor da totalidade do texto apresentado, não aprento texto plagiado, e tomei conhecimento das consequências de uma situação de plágio.

Porto, 29 de Setembro de 2017

Acknowledgments

The present work was elaborated and performed in the Phenotypic Evolution labs of the i3s building, as part of the dissertation thesis of the Cell and Molecular Biology Master's Degree, part of the Faculty of Sciences of the University of Porto. To whom I directed my acknowledgment and thanks. My sincere thanks to the contributions of many individuals without whom this work wouldn't be possible.

To FCUP, i3s, IBMC and FCT for the opportunity given to perform this work, the facilities and conditions provided.

To Jorge Vieira and Cristina Vieira, my advisor and co advisor, for taking the chance with an unknown student to work on one of their many interesting projects, to whom I thank the vote of confidence, the adquired experience and the valuable knowledge obtained.

To my lab mates, Vanessa, Rodrigo, Sara and Pedro for the shared experiences, the knowledge sharing and the good welcoming in the group.

To my mom, dad and sisters, to whom I thank the possibility of providing for my education, in my master's degree, and the support given in all stages of my life.

To my grandmothers for the support and interested shown in this last 2 years.

To my friends and collegues, Necati, Rui, Marina, Sara, Carolina, Lissa, Eliana, Miguel, Cláudio for the support given, the shared moments of friendships, the shared experiences, and for the many years ahead.

To every person who has in someway contributed to this present project, my sincere thanks.

Abstract

Development time differences can have an impact on the organism and its history. In *Drosophila* there is an association between development time, lifespan and adult size. *Drosophila* is an ideal model for development studies since genomic data is available for many species, including *D. americana* and *D. novamexicana*, furthermore, developmental time can be easily phenotyped as here performed, it is important to understand the molecular basis of development time differences. A bioinformatics search for candidate genes was performed, focusing on genes with signs of positive selection, according to the bpositive database, furthermore, location on Muller's element C was also considered, since an association between this chromosomal arm and developmental time [1] was seen reported in association with development processes and/or signaling pathways. Six candidate genes were chosen (*Sin3A, sm, CG3209, yki, hpo,* and *Act79B*). The strains used (*D. americana* strain W11; *D. americana* strain SF12; *D. novamexicana* strain 15010-1031.00, here on mentioned as NOVA00*),* belong to the *virilis* group of species, a group with special interest because of the wide range of phenotypic differences found between closely related species. The expression of the candidate genes was studied using RT-PCR under 12 different conditions of development time and days of life. For *Sin3A, sm, CG3209, hpo,* and *yki* there are no significant differences in expression between flies with different development times at the same day of life, and for flies at different days of life with the same development time. For *Act79B* there are no significant differences between the flies with different development times at the same day of life. There is, however, a significant difference in gene expression between flies at different days of life and the same development time (the expression of *Act79B* decreases with age). Taking into account the findings for *Act79B* gene expression for different development times at the same day of life we can argue that it is not unexpected that the candidate genes are not showing differences in expression in flies with different development times, as *Act79B* expression levels suggests that there isn't an ontogenetic development difference in flies with different development times at 4 and 8 days of adult life. It is unclear what causes a fly to have different days of development (15-16-17), since none of the selected genes could explain the difference between development times. Moreover, according to our findings, we hypothesize that development time variations and ontogenetic development of adult in *Drosophila* are not correlated.

KEY-WORDS: *D. americana*; *D. novamexicana*; Development Time; Positive Selection; Gene Expression; Candidate Genes; Phenotype-Genotype

Resumo

Em *Drosophila*, existe uma associação entre tempo de desenvolvimento, longevidade e tamanho dos adultos. *Drosophila* é um modelo ideal para estudos de desenvolvimento, uma vez que informações genómicas estão disponíveis para a maior parte das espécies, incluindo *D. americana* e *D. novamexicana*, e o tempo de desenvolvimento pode ser facilmente fenotipado, é importante entender quais as bases moleculares que explicam as diferenças no tempo de desenvolvimento. A pesquisa bioinformática de genes candidatos focou-se na procura de genes com sinais de seleção positiva, de acordo com a base de dados bpositive, para além da presença no elemento Muller C, uma vez que existe uma associação entre este braço cromosomal e o tempo de desenvolvimento [1], relevando associação com outros processos relacionados com o desenvolvimento e/ou vias de sinalização. Seis genes foram escolhidos (*Sin3A, sm, CG3209, yki, hpo,* e *Act79B*). As estirpes estudadas (W11 de *D. americana*; SF12 de *D. americana* e 15010-1031.00 de *D. novamexicana*, daqui em diante mencionada como NOVA00*)*, pertencem ao grupo de espécies *virilis*, um grupo com especial interesse devido à ampla variação fenotípica presente entre espécies proximamente relacionadas. A expressão destes genes foi estudada através de RT-PCR em 12 condições diferentes de tempo de desenvolvimento e dias de vida. Para *Sin3A, sm, CG3209, hpo,* e *yki*, não há diferenças significativas de expressão entre moscas com diferentes tempos de desenvolvimento no mesmo dia de vida e para moscas com diferentes dias de vida com o mesmo tempo de desenvolvimento. *Act79B*, não apresenta diferenças significativas de expressão em moscas com diferentes tempos de desenvolvimento no mesmo dia de vida. No entanto, existe uma diferença significativa na expressão de *Act79B* em moscas em dias diferentes de vida com o mesmo tempo de desenvolvimento (a expressão de *Act79B* decresce com a idade). Considerando os resultados para a expressão de *Act79B* em moscas com diferentes tempos de desenvolvimento e no mesmo dia de via, podemos sugerir que não é inesperado que os genes candidatos não mostrem diferenças de expressão em moscas com diferente tempo de desenvolvimento, uma vez que os níves de expressão de *Act79B* sugerem que não existe uma diferença no desenvolvimento ontogenético em moscas com maior tempo de desenvolvimento aos 4 e 8 dias de vida adulta. É questionável o que causa *Drosophila* a ter diferentes tempos de desenvolvimento (15-16-17), uma vez que nenhum dos genes selecionados pôde explicar estas diferenças. De acordo com os nossos resultados, sugerimos que a variação no tempo de desenvolvimento e o desenvolvimento ontogenético de adultos em *Drosophila* não estão correlacionados.

PALAVRAS-CHAVE: *D. americana*; *D. novamexicana*; Tempo de Desenvolvimento; Seleção Positiva; Expressão Genética; Genes Candidatos; Fenótipo-Genótipo

Index

Index of Figures

Index of Tables

List of Abbreviations

ΔCT	Variation of cycle threshold
AS	Abdominal Size
BDBM	BLAST Database Manager
CDS	coding sequences
CRT	Chill-Coma Recovery Time
CT	Cycle Threshold
DT	Development Time
Egfr	Epidermal growth factor receptor
GO	Gene Ontology
Grs	Gustatory Receptors
GST	Glutathione S-Transferase
LS	LifeSpan
miRNA	microRNA
mtDNA	mitochondrial DNA
My	Million Years
ncRNA	non-coding RNA
Ors	Olfactory Receptors
P450	cytochrome P450
RNAi	RNA interference
RQI	RNA quality indicator
RT	Room Temperature
snRNA	small nuclear RNA
TAE	Tris-Acetate-EDTA
Tm	melting temperature
tRNA	transfer RNA

I. Introduction

I.1. Project Presentation and Objectives

Development time differences can have an impact on the organism and its history, being widely associated with adaptation of species. Moreover, developmental time shows an association with lifespan and adult size among other traits [1]. Thus, it is important to understand the molecular basis of development time differences. One of the approaches to address this question is to search for genes that show variation at the amino acid level and that might affect the phenotypic differences observed in terms of developmental time. Another one is to compare the gene expression of candidate genes in individuals showing different developmental times. The model organism *Drosophila* is an ideal model for such studies since genomic data is available for many species, they are easy to maintain, and developmental time can be easily phenotyped.

Development has been well studied in *Drosophila* species and several studies show that for development time the type of food is one of the biggest modifiers of the phenotype [2]. Several other studies have made genetic and nutritional manipulations with the objective of observing the effects of this in development time of *Drosophila*, showing several degrees of association between the studied manipulations and the shown phenotypes [3]. However, in the present work, the goal was to study the observed differences in the development time of *Drosophila*, without subjecting flies to any genetic or nutritional manipulation.

This project focused on the study of the expression of candidate genes who could in principle explain the differences in development time in 3 strains of *Drosophila* (*D. americana* strain W11; *D. americana* strain SF12; *D. novamexicana* strain NOVA00), all belonging to the *virilis* group. These species group is interesting in many areas of studies, because of the wide range of phenotypic differences found between closely related species.

Genes on Muller's element C were chosen because of the association between this chromosomal arm and developmental time [1]. The expression of the candidate genes was studied using RT-PCR under 12 different conditions, namely, strain W11 with 15 development days at 4 days of life and at 8 days of life; W11 with 16 development days at 4 days of life and at 8 days of life; strain SF12 with 15 development days at 4 days of life and at 8 days of life; SF12 with 16 development days at 4 days of life and at 8 days of life; strain NOVA00 with 16 development days at 4 days of life and at 8 days of life; NOVA00 with 17 development days at 4 days of life and at 8 days of life.

In this work, firstly, we present a review of the knowledge on the characteristics that make *Drosophila* an ideal model organism to study many biological processes, including development.

We present a short review of its importance in the development of scientific knowledge around many areas of interest, the importance of the sequencing of the original 12 *Drosophila* genomes and their features, the main characteristics of *Drosophila* genus and the reason why *virilis* group is a remarkably interesting group of species. The comparisons between *D. americana* and *D. novamexicana*, the two *Drosophila* species used in this project are discussed. Then, we briefly review the relevant literature in development. Moreover, we present evidence of positive selection present in *Drosophila* genes involved with DT (development time). Lastly, the main results are discussed and their contributions to the topic of interest.

I.2. The Known Potential of *Drosophila* as a Model Organism in Biological Studies

Many important features of *Drosophila*, show its great potential in molecular and cellular biology studies. Some of the biggest advantages of using the fruit fly are its short life cycle, easiness to keep in laboratory conditions, easiness of hybridization with closely related species and the numerous different strains with a wide range of phenotypic traits, among several other interesting characteristics.

The rapid and powerful genetic tools available for *Drosophila* in recent years have maximized the potential of the fruit fly in biology, elucidating gene mechanisms, and revealing novel disease-relevant biology [4]. Information on these resources and many other useful genetic reagents are available through flybase.org, an online annotated resource for *Drosophila* genetics (http://flybase.org) [5, 6].

The *Drosophila* system has also been used as a model to develop bioinformatic tools for the characterization of coding and regulatory regions [7, 8].

Considering the substantial genomic conservation with humans, it is no surprise that studies in *Drosophila* have touched almost all branches of human disease biology, evidencing the ever-growing importance of *Drosophila* as an organism model [9]. Many different studies have produced interesting results touching several different human diseases; this includes studies of sterol absorption in the gut [10], neuromuscular dysfunction in mucolipidosis [11], mechanisms of congenital renal disease [12] and cardiotoxicity from a high-fat diet [13]. The several *Drosophila* models have warranted powerful insights into several neurodegenerative disorders [14, 15, 16] such as Alzheimer's and Parkinson's disease [15], amyotrophic lateral sclerosis [17], Huntington's disease [18, 19] and spinocerebellar ataxias [20]. Mechanisms to identify conserved regulators of pain perception [21], cardiac function [22] and adipocyte differentiation [23], have produced further insights into human disease. Many other screens have highlighted Mendelian disease genes [24].

It should be noted that all these studies rely on the key assumption that homologous genes between *Drosophila* and human will have conserved functional activities, leading to similar phenotypes when subject to any genetic manipulation besides the assumption of evolutionary conservation at the sequence level [4].

I.3. The *Drosophila* Genus

I.3.1. The Importance of the Drosophila Genomes Sequencing Project and its Results

Although *D. melanogaster* is a lot more extensively characterized and studied than other *Drosophila* species, in recent years, other species of the genus have been the subject of many studies. One important milestone in *Drosophila* biology was the sequencing of 12 *Drosophila* species (*D. melanogaster*, *D. sechellia*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis and D. grimshawi*) (Fig. 1), known widely as the *Drosophila* 12 Genomes Consortium [25]. Important advances were made regarding the history, evolution, and the further understanding of the divergence of traits between species. The genus *Drosophila* includes both closely related species and species that are diverging for at least 40 My (million years) [26, 27]. For instance, *D. melanogaster*, *D. americana and D. novamexicana* (the latter two of the *virilis* group and the subject species of our study) are diverging for at least 40 My [26, 27]. Many other species have been sequenced since that time, including *D. americana.*

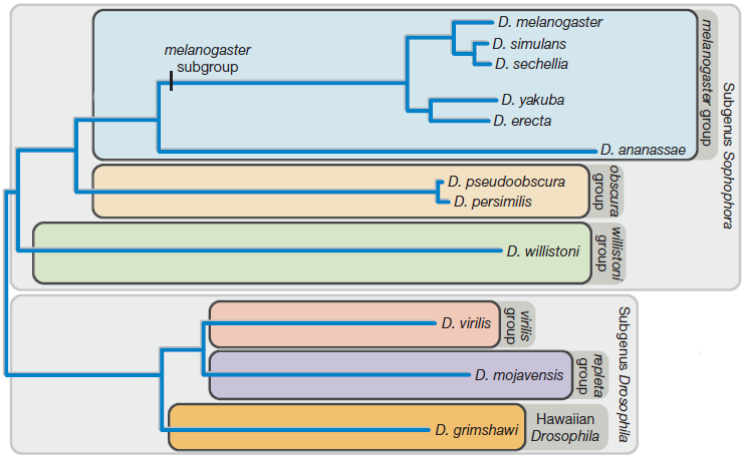


Figure 1 - Phylogram of the 12 sequenced species of *Drosophila*, derived using pairwise genomic mutation distances and the neighbor-joining method (adapted from [25])

It is established that the most important aspects of the cellular, molecular and developmental biology of the 12 sequenced species are well conserved. Therefore, the genomes of these species provide an excellent model for studying how conserved functions are maintained in the face of sequence divergence, as well as providing extensive resources for the study of the relationship between genome sequence and phenotypic diversity. The sequences provide a fantastic tool to contrast genome structure, genome content, and evolutionary dynamics across the well-defined phylogeny of the sequenced species (Fig. 1) [25].

As mentioned above, at a broad level, the overall genome structure is well conserved across the 12 sequenced species. This can be seen in several characteristics presented on Table1; here we highlight the differences between the 12 species, focusing on *D. melanogaster*, the model organism, and *D. virilis*, the closest species to *D. americana* and *D. novamexicana*. Total genome size (estimated by flow cytometry) has small variation ranging from 130Mb (*D. mojavensis*), 200Mb in *D. melanogaster* and further 364Mb (*D. virilis*) (Table 1), in contrast with a reported difference between *Drosophila* and mammals [28]. Total protein-coding sequence ranges from 38.9Mb in *D. melanogaster*, 57.9Mb in *D. virilis*, to 65.4Mb in *D. willistoni*. Intronic DNA content ranges from 19.6Mb in *D. simulans*, 21.7Mb in *D. virilis*, 21.8Mb in *D. melanogaster* to 24.0 Mb in *D. pseudoobscura* (Table 1). Despite the conservation found in this characteristics, there are many putatively non-neutral changes identified in protein-coding genes, non-coding RNA genes, and cis-regulatory regions. These differences may prove to be the basis of divergences in the ecology and behavior of these species [25]. Approximately 20% of transcription in *D. melanogaster* seems to be unassociated with protein-coding genes [29] and between 5.6–32.3% of gene models in non-*melanogaster* species correspond to protein-coding derived from transposable elements. There were 8,563 genes with single-copy orthologous genes in the *melanogaster* group and 6,698 genes with single-copy orthologous genes in all 12 species (Figure 2) [25, 30]. From the 21,928 cases in which a *D.melanogaster* gene was absent from another species, there were further identified plausible homologs for 13,265 (60.5%), confirmed absences of homology for 4,546 (20.7%) genes and the remaining 4,117 (18.8%) were unable to be resolved [25]. The number of ncRNA (non-coding RNA) genes per family in *Drosophila* is relatively low (Table 1). tRNAs (transfer RNAs) genes are the most abundant family of ncRNA genes in all 12 genomes, with 297 tRNAs in *D. melanogaster*, 279 in *D. virilis* and between 261–484 tRNA genes in the other species (Table 1). From the 78 previously reported miRNA (microRNAs) genes, 71 (91%) are highly conserved across the entire genus, with the remaining seven genes restricted to the subgenus *Sophophora*. All species contain identical numbers of snRNA (small nuclear RNA) genes (Table 1)[25].

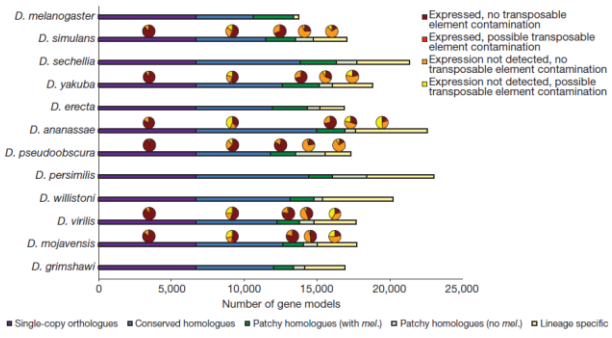


Figure 2- Number of gene models in 12 *Drosophila* genomes that fall into one of the five mentioned homology classes (adapted from [25], see for further details).

Table 1- A summary of annotated features across all 12 sequenced genomes (adapted from [25]).

	Protein-coding gene annotations		tRNA (pseudo)	Non-coding RNA annotations				Repeat coverage (%)	Genome size (Mb; assembly /flow cytometry)
	Total no. of protein- coding genes (per cent with *D. melanogaster* homologue)	Coding sequence/ intron (Mb)		snoRNA	miRNA	rRNA (5.8S + 5S)	snRNA		
D. melanogaster	13,733 (100%)	38.9/21.8	297 (4)	250	78	101	28	5.35	118/200
D. simulans	15,983 (80.0%)	45.8/19.6	268 (2)	246	70	72	32	2.73	111/162
D. sechellia	16,884 (81.2%)	47.9/21.9	312 (13)	242	78	133	30	3.67	115/171
D. yakuba	16,423 (82.5%)	50.8/22.9	380 (52)	255	80	55	37	12.04	127/190
D. erecta	15,324 (86.4%)	49.1/22.0	286 (2)	252	81	101	38	6.97	134/135
D. ananassae	15,276 (83.0%)	57.3/22.3	472 (165)	194	76	134	29	24.93	176/217
D. pseudoobscura	16,363 (78.2%)	49.7/24.0	295 (1)	203	73	55	31	2.76	127/193
D. persimilis	17,325 (72.6%)	54.0/21.9	306 (1)	199	75	80	31	8.47	138/193
D. willistoni	15,816 (78.8%)	65.4/23.5	484 (164)	216	77	76	37	15.57	187/222
D. virilis	14,680 (82.7%)	57.9/21.7	279 (2)	165	74	294	31	13.96	172/364
D. mojavensis	14,849 (80.8%)	57.8/21.9	267 (3)	139	71	74	30	8.92	161/130
D. grimshawi	15,270 (81.3%)	54.9/22.5	261 (1)	154	82	70	32	2.84	138/231

Similar numbers of predicted protein-coding (Table 1) genes are observed across the 12 species. The majority of predicted genes in each species have homologs in *D. melanogaster* (Table 1). Most of the 13,733 protein-coding genes in *D. melanogaster* are conserved across the entire phylogeny: 77% have identifiable homologues in all 12 genomes, 62% can be identified as single-copy orthologues in the six genomes of the *melanogaster* group and 49% can be identified as single-copy orthologues in all 12 genomes (Fig. 2). Figure 2 summarizes studies by [25], presenting a number of gene models in 12 *Drosophila* genomes that fall into one of five homology classes: single-copy orthologues in all species (single-copy orthologues), conserved in all species as orthologues or paralogues (conserved homologues), a *D. melanogaster* homologue, but not found in all species (patchy homologues with *mel.*), conserved in at least two species but without a *D. melanogaster* homologue (patchy homologues, no *mel.*), and found only in a single lineage (lineage-specific). Furthermore, for species with expression data available, the fraction of genes in each homology class that fall into one of four evidence classes is presented. The vast majority (11,348/11,644, or 97.5%) of *D. melanogaster* proteins that can be unambiguously assigned a homology pattern are inferred to be ancestrally present at the genus root. Of the remaining 296 non-ancestrally present genes, 252 are *Sophophora*-specific. The remaining 44 proteins include 14 present in the *melanogaster* group, 23 present only in the *melanogaster* subgroup, 3 unique to the *melanogaster* species complex, and 4 found in *D. melanogaster* only [25].

The vast majority of multigene families are found in all 12 genomes, however, their size is highly dynamic: 4,692 (41.0%) gene families changed in size in at least one species, and an appreciable fraction shows rapid and lineage-specific expansions and contractions. There are at least 44 protein-coding genes unique to the *melanogaster* group, proteins who were found to have very different properties from ancestral ones [25].

Across the genus, transposable elements range from 2.7% in *D. simulans* and *D. grimshawi*, 5.35% in *D.melanogaster*, 13.96% in *D.virilis* to 25% in *D. ananassae* (Table 1). Within the *Drosophila* subgenus, *D. grimshawi* has the lowest transposable element content, possibly relating to its ecological status as an island endemic, which potentially minimizes the chances for horizontal transfer of transposable element families. The highest levels of

transposable elements are found in *D. ananassae*, who also has the highest numbers of pseudo-transfer (t)RNA genes (Table 1). These results further suggest a potential relationship between pseudo-tRNA genes and repetitive DNA, in accordance with what has been established in the mouse genome [31].

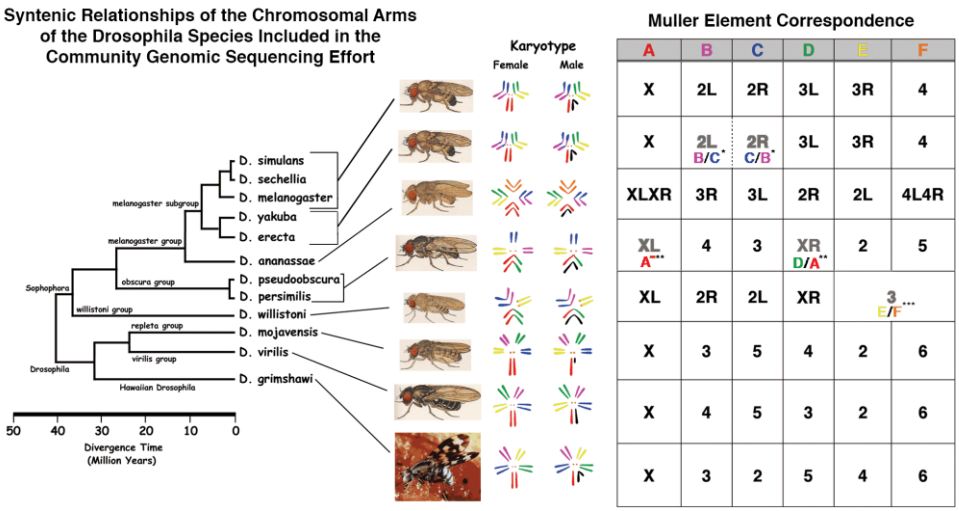
When looking into specific genes and parts of the genome, the number of structural changes and rearrangements is obviously larger; one interesting case is the several different rearrangements of genes in the Hox cluster found in *Drosophila* species [25]. Several cluster splits have been identified in *Drosophila* [32, 33] and the 12 *Drosophila* genome sequences provide additional evidence of Hox gene clustering in *Drosophila*. There are six different gene arrangements found, with no species retaining the inferred ancestral gene order [34].

I.3.2. The Genus Main Characteristics

The *Drosophila* genus is a large, diverse and widely distributed group of organisms [35]. Its taxonomy is relatively well known and established, there is broad information for most of the species [37]. The genus is paraphyletic as several other genera are included within the phylogeny of *Drosophila* [36]. Ten subgenera are presently recognized within the genus, of which *Sophophora* and *Drosophila* are the major ones [38]; these are further taxonomically subdivided into radiations and species groups. Of the two main subgenera, *Drosophila* is the largest one, comprising three major lineages: the *funebris* species group; the *virilis-repleta* radiation; and the *immigrans-tripunctata* radiation.

The origin of the *virilis-repleta* radiation has been placed in the Old World tropics [39], most likely in the Asiatic tropical regions [35]; from this ancestor two lineages evolved leading to the Old World tropics (e.g. the *polychaeta* group) and temperate species groups (e.g. *virilis*, *robusta* and *melanica* species groups). A Neotropical radiation, which comprises the *repleta*, *canalinea*, *mesophragmatica*, *dreyfusi*, *annulimana* and *nannoptera* species groups, evolved from a third lineage of the Asiatic tropical ancestor.

The crown age for the *Drosophila* subgenus (and, therefore, the divergence of the two major lineages, the *immigrans-tripunctata* and *virilis-repleta*), is placed in the late Eocene, approximately 34 Mya [27, 40]. Other studies indicate that the major groups appeared during the Oligocene/Miocene transition or early Miocene [27].



A	B	C	D	E	F
X	2L	2R	3L	3R	4
X	2L B/C*	2R C/B*	3L	3R	4
XLXR	3R	3L	2R	2L	4L4R
XL A***	4	3	XR D/A**	2	5
XL	2R	2L	XR	3 ... E/F***	
X	3	5	4	2	6
X	4	5	3	2	6
X	3	2	5	4	6

Figure 3- Syntenic Relationships of the Chromosomal Arms of the *Drosophila* Species (adapted from flybase.org [6])

Furthermore, the genus *Drosophila* varies considerably in their morphology, ecology, and behavior [41]. From an ecological point of view, the members of the subgenus *Drosophila* occupy a wide variety of niches, from sap feeders (e.g. *robusta*, *melanica* and *virilis*) to cactophilic species (e.g. *repleta*), mycophagous (e.g. *quinaria*) and flower feeders (e.g. *tripunctata*) [37, 38], although many of the species are generalists and can exploit different resources.

Colonization of new geographic regions and new ecological resources can result in rapid species diversification into the new ecological niches available, a phenomenon that seems to have shaped the evolution of *Drosophila* genus. Members of the subgenus *Drosophila* are distributed across the globe [27], spanning a wide range of global distributions, the 12 original sequenced species [25] originate from Africa, Asia, the Americas, the Pacific Islands and also cosmopolitan species have colonized the planet (*D. melanogaster and D. simulans*) as well as closely related species that live on single islands (*D. sechellia*) [42]. A variety of behavioral strategies is also encompassed by the sequenced species, ranging in feeding habit from generalist such as *D. ananassae*, to specialist such as *D. sechellia*, which feeds on the fruit of a single plant species.

Furthermore, *Drosophila* species vary in their number of chromosomes and arrangements. There are six fundamental chromosome arms common to all species, referred to as 'Muller elements', and named from A to F. Although most pairs of orthologous genes are found on the same Muller element, there is gene shuffling within Muller elements between diverged genomes (Fig. 3). Moreover, inversions seem to have played important roles in the process of speciation in at least some taxa [43].

Despite this wealth of diversity, *Drosophila* species share a distinctive body plan and life cycle [25].

I.4. The *virilis* Group of Species

I.4.1. virilis Group Main Characteristics

The *virilis* group of Drosophila is of special interest because it comprises many closely related species with different phenotypic traits evolving for a relatively short period. The species group provides a good model for analyzing the genetic basis of morphological differences [26, 44, 45] being one of the best-studied species groups in the subgenus *Drosophila*.

In total, the *virilis* group comprises 12 closely related species, separated in two main phylads, *montana* and *virilis*. The *virilis* phylad—includes *D. virilis, D. lummei, D. novamexicana,* and *D. americana* [26]. There are distinct differences in chromosome number and shape, overall ecology, morphology, and geographic range. The group has a wide geographic range stretching from the tropics to the northern taiga in both Old and New Worlds, comprising species living in deserts, in the tropics, on chains of volcanic islands and commensally with humans. Larvae of these fruit flies develop in rotting wood, mainly that of deciduous trees of the genus *Salix*. Gross morphological differences among these taxa are minimal, except for differences in the structure of the male genitalia [46]. The group exhibits a series of chromosomal rearrangements, and the resulting differences among karyotypes were primary characters originally used to define taxa within the group [47]. Chromosomal changes consisting of centromeric fusions and inversions are prevalent. Several chromosomal forms such as *D. novamexicana* and *D. americana* apparently arose through dispersal of an ancestral lineage into North America [47-49].

One striking difference in the *virilis* group of species is the pupal case color [50]. All members of the group have brown puparia, except for *D. virilis* who has black puparia that are easily distinguished in early pupal development. Furthermore, this trait is not entirely fixed across the species, the intensity of the black color varies between different strains and is affected by crowding and culture conditions [50-52]. A single genomic interval (11 kb) on chromosome 5 was identified as the cause of the pupal case color difference. This region contains only the first exon and regulatory region of a single *D. virilis* gene, *GJ20215*, which is the homolog of the *D. melanogaster* dopamine N-acetyltransferase (*Dat*), an enzyme known to act in the pigmentation pathway. It seems that reduced expression of *Dat* in early pupal development in *D. virilis* is the cause of dark pupae in this species [52]. *Dat* is among a few other genes identified in *Drosophila* (e.g. [53-55]) that by themselves can cause striking phenotypic differences in closely related species [52].

Regardless of the observed differences, many species can cross with one another at least in one direction and produce partly fertile offspring [56].

I.4.2. Phylogenetic Relationships in the virilis Genus

Several studies have examined the taxonomy, phylogenetic history and reproductive incompatibilities among members of this group [47, 52, 57-69].

It seems that interspecific hybridization has significantly contributed to the evolutionary history of the *virilis* species group [56].

Using a set of 48 polymorphic microsatellites derived from *D. virilis* to infer phylogenetic relationships in the *virilis* clade, the results were consistent with previous studies, showing *D. virilis* and *D. lummei* as the most basal species of the group [70]. *D. lummei* is sister to the North American members of this subgroup. Both *D. virilis* and *D. lummei* are Palearctic endemics [64]. *D. virilis* being endemic to middle latitudes in Asia and *D. lummei* occurring at higher latitudes [64]. Moreover, the division between *D. lummei* and the North American complex of *D. novamexicana* and *D. americana* apparently represents the dispersal of the *virilis* subgroup into the Nearctic region. Fly lines of the *virilis* subgroup originating from North America, *D. novamexicana* and *D. americana*, form a strongly supported monophyletic group of closely related sequences [47]. The emergence of an ancestral lineage in North America through a common ancestor with *D. lummei* is consistent with dispersal through a Beringian corridor. Several chromosomal inversions originated in the common ancestor are currently polymorphic in both *D. lummei* and *D. americana* [64], which suggests their ancestral population was well connected and that both lineages have maintained historically large effective sizes [47].

I.4.3. D. virilis

Besides pupal case color, *D. virilis* differs from other species of the *virilis* group in that inversion chromosome polymorphism is absent in its natural populations. Analysis of the nucleotide sequences of different gene families invariably leads to the conclusion that *D. virilis* is an ancient species which was one of the first to diverge from the common ancestor of the group several million years ago [56, 68, 71-72]. The hypothetical place of origin of *D. virilis* is in Southeast Asia [56, 73].

I.4.4. D. virilis and D. americana Relationship

D. virilis and *D. americana*, are two species of special interest because, among other reasons, they can be crossed easily and yield hybrids which are partially fertile [74]**.** *D. virilis* has been diverging from *D. americana* for approximately 4.1 My [26, 60]. This two sub-species differ in gene alignment, in the chromosome configuration of the ganglion cells, in phenotypic appearance, and physiological characteristics. *D. americana* has a much larger eye, with finer eye pile, broader carina, overall darker body color, heavier cloud on posterior cross-vein and a more fusiform body [75].

I.4.5. D. americana

D. americana is a member of the *virilis* subgroup of *Drosophila* and is closely related to *D, virilis* [68]. *D. americana* and *D. melanogaster* have been diverging for about 40My [1, 27]. 70.2% of the annotated *D. melanogaster* proteins have a hit in *D. americana* genomes (15,308 hits; Table 1). When using the same approach, 84.1% of the annotated *D. melanogaster* proteins have a hit in the *D. virilis* genome (18,934 hits; Table 1), but this likely reflects differences in genome coverage [76]. Its ecology means that it is less likely to have been disturbed by human activities than species such as *D. melanogaster* and *D. simulans* [77]. *D. americana* is native to the United States where it has been independently evolving for approximately 1 Myr [26, 47, 78]. This species is widely distributed, across the Central and Eastern regions of the United States from the South (Texas to the states around the Gulf of Mexico) to the North of the country (from Montana to Maine) [73]. This species can be easily collected along the margins of marshes, lakes, and rivers, especially those where there is a high density of *Salix* species [64], and in recent years, several articles were published using hundreds of wild-caught *D. americana* individuals from different populations and strains [26, 79-82].

D. americana is an excellent species for comparative studies on the molecular basis of phenotypic variation [1, 26]. Besides *D. virilis*, this species can also be crossed with *D. novamexicana*. *D. americana* is thought to present a large amount of genotypic variation, low levels of population structure and a stable historical population size [26, 78, 83]. The inferred *D. americana* effective population size is 3.6 million individuals [26].

I.4.6. D. novamexicana and D. americana Relationship

D. novamexicana is another member of the *virilis* group of species, that is very closely related with *D. americana*, however, they appear to be different in their evolutionary histories. Results show that *D. americana* has had a large, stable population and there is no evidence that indicates any recent significant reduction in population size. Conversely, there is evidence in the results for a reduced ancestral population size in *D. novamexicana*. The findings are consistent with a recent cladogenesis event of *D. novamexicana* and *D. americana*. Furthermore, *D. novamexicana* as a population is significantly differentiated relative to *D. americana* [66, 84]. *D. americana* inhabiting the Central and Eastern regions from the South (Texas to the states around the Gulf of Mexico) (Fig.4) to the North of the country (from Montana to Maine), while *D. novamexicana* has a more restricted distribution in the Southwest (New Mexico, Arizona, Colorado and Eastern Utah) (Fig.5) [73]. The age of divergence between these two species was estimated to be 0.866 My and 0.266 My for *fused1* and *lim3*, respectively, with one method. Estimated 0.388 My with Stephens et al. (1985) [85] method and 0.380 My by Caletka et al. (2004) [47]. Furthermore, other lines of evidence support the recent age of these species such as the estimated age for the inversion *Xc*, 0.308 My, Vieira *et al.* (2006) [86]. The divergence time seems to be around 0.40 My, considering the several lines of work and different approaches taken into aging the divergence of this two species. Maintenance of the ancestral karyotype in *D. novamexicana* contrasts distinctly with the occurrence of two centromeric fusions in *D. americana*. The restricted geographical location of *D. novamexicana* to the western slopes of the Rocky Mountains suggests that *D. novamexicana* evolved from a marginal, peripheral population of the *D. americana*/*D. novamexicana* ancestral species. The mid to late-Pleistocene is a period characterized by pluvial–interpluvial cycles in Southwest North America that have influenced the evolution and diversification of many North American species [87-88]. Thus, the effect of the pluvial–interpluvial period of the Pleistocene appears to be restricted to the most Southwestern population of the ancestral species. It is also evident that *D. novamexicana* evolved in complete isolation from *D. americana* since there is no conclusive evidence for introgression of *D. americana* alleles [89].

Figure 4- Geographical location of *D. americana*, found east of the Rocky Mountains, a major mountain range in western North America (adapted from [90])

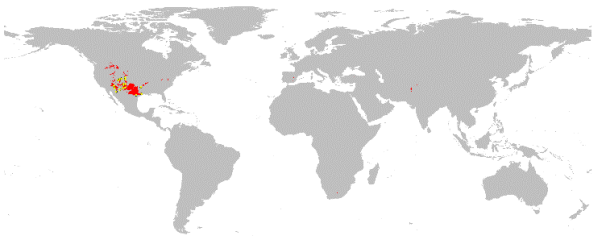


Figure 5 - Geographical location of *D. novamexicana*, found in sparse populations confined to the southwestern desert of the United States, around Arizona and New Mexico (adapted from [90])

The two species differ in their karyotype; *D. americana* has a fixed fusion of chromosomes 2 and 3 while *D. novamexicana* retained the ancestral karyotype where chromosomes 2 and 3 are unfused [75, 91]. There are also further karyotype differences between the two species in the frequency of some polymorphic chromosomal inversions. Interestingly, there are populations of *D. americana* that have chromosomal rearrangements in common with *D. novamexicana* while other populations of *D. americana* show the alternative arrangement [1, 92-94].

However, one of the most easily recognizable differences between *D. americana* and *D. novamexicana* is the pigmentation pattern. It is striking how such closely related species differ so greatly on such an important phenotypic trait. Adults of *D. novamexicana* differ in color from their sister species as they have evolved a light brown color pigmentation along the dorsal abdomen, head, and thorax, while all other members of the group are darkly pigmented [44, 93]. Furthermore, *D. novamexicana* also lacks pigment along the abdominal dorsal midline (Fig.6) [93, 95].

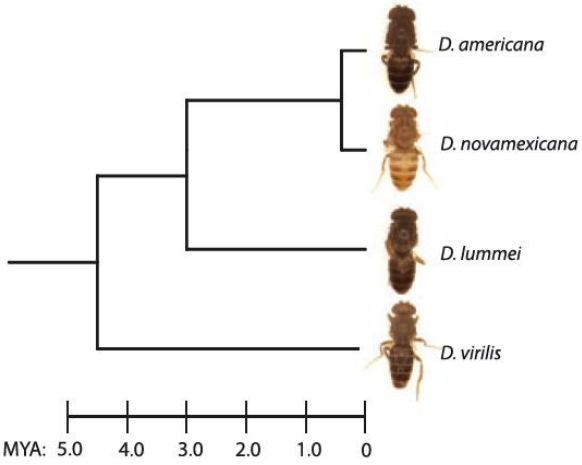


Figure 6 - Pigmentation differences in the *virilis* group of *Drosophila*. Estimated divergence times [47] are shown on the scale at the bottom (MY=millions years) (adapted from [95]).

Pigmentation differences in adult insects have been attributed to several selective pressures, including desiccation resistance, ultraviolet protection, thermal regulation, crypsis and sexual

selection [45, 96-100]. Selection on pigmentation phenotypes may also be indirectly caused by pleiotropic genes in the pigmentation pathway, affecting a number of traits [101-102]. Previous works have shown that *ebony* plays a major role in the pigment difference [93] and that *tan* also contributes [102], the two genes account for nearly 87% of the pigmentation difference [95, 104]. Abdominal expression of the *Ebony* protein is higher in *D. novamexicana* than in the darker *D.americana*, consistent with its ability to inhibit melanization [93].

Furthermore, pigmentation has emerged as a model trait for evolutionary and developmental analysis that is particularly amenable to molecular investigation in the genus *Drosophila* [103], because this genus combines the advantages of a well-developed model system with substantial phenotypic diversity. As previously mentioned, the availability of genetic tools in *D. melanogaster* has increasingly helped elucidate several genetic and developmental mechanisms of pigment evolution [95]. Many studies have touched this topic [93, 95-96, 99, 104-109].

I.5. Evolution in the *Drosophila* Genus

I.5.1. Overall Evolution

As one of the main goals of this project was to study genes showing signs of positive selection according to the bpositive online database [110] (http://bpositive.i3s.up.pt/), while also being involved in development, it is of special importance to understand how the genome is evolving in the *Drosophila* genus. Many studies have approached the evolutionary processes in *Drosophila* genomes across the phylogeny of species. The *Drosophila* 12 Genomes Consortium [25] has collected several analyses on how are the genomes and sequences evolving in the genus. Foremost, having multiple genomes in a phylogenetic framework improves dramatically the precision and sensitivity of evolutionary inference through comparative analysis, producing more robust results than single-genome analyses. The data gathered from the sequencing of the 12 *Drosophila* species genomes presented by the Drosophila 12 Genomes Consortium [25] illustrate how rates and patterns of sequence divergence across taxa can enlighten evolutionary processes on a genomic scale. The several sequenced genomes of the *Drosophila* genus provide great statistical power to identify factors affecting rates of protein evolution.

Although the level of gene expression consistently seems to be a major determinant of variation in rates of evolution among proteins [111-112], studies have suggested that there are other factors playing a significant part [113-116]. The breadth of expression across tissues, gene essentiality, and intron number all independently correlate with rates of protein evolution, suggesting this are also important factors in modulating rates of protein evolution [117]. The presence of repetitive amino acid sequences seems to play a role too, non-repeat regions in proteins containing repeats evolve faster and show more evidence for positive selection than genes lacking repeats [118].

The evolutionary divergence spanned by the genus *Drosophila* exceeds that of the entire mammalian radiation when generation time is taken into account [25, 119].

Sackton et al. (2007) [120] studied 226 genes using data from the 12 sequenced *Drosophila* species genomes and about 10% of the genes analyzed showed evidence of positive selection, using a likelihood-based framework [121].

I.5.2. GO Terms Evolution

In the *Drosophila* 12 Genomes Consortium study [25], several GO (Gene Ontology) terms were evaluated for rates of gain/loss. The most common terms with elevated rates of gain/loss include "defense response", "protein binding", "zinc ion binding", "proteolysis", and "trypsin activity". Genes present in GO terms "defense response" and "proteolysis" also show high rates of protein evolution. Overall heterogeneity was observed in rates of gene gain and loss across lineages, suggesting that there is no lineage with an increase/decrease in gene gain and loss [122].

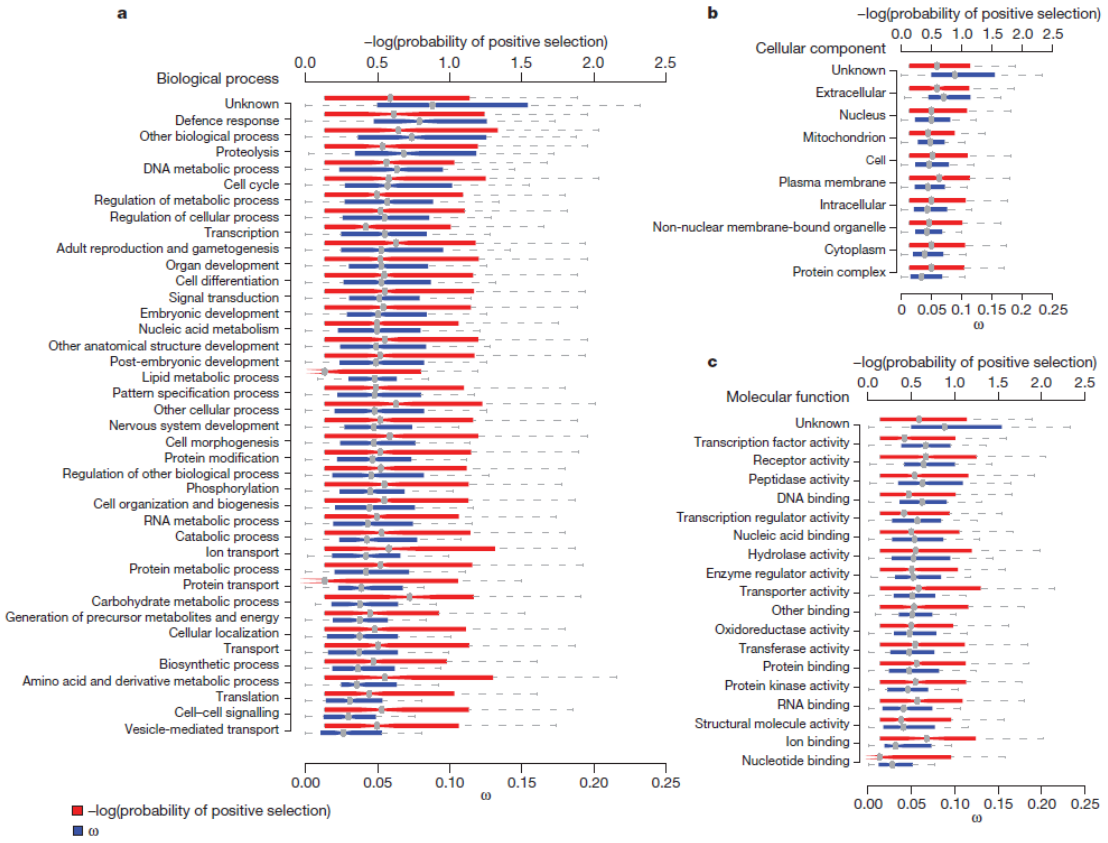


Figure 7- Patterns of constraint and positive selection among GO terms: a, biological process GO terms; b, cellular component GO terms; and c, molecular function GO term. Distribution of average ω per gene is shown and the negative log10 of the probability of positive selection for the selected genes (adapted from [25]).

Most of the functional categories of genes are strongly constrained, with median estimates of ω=dN/dS (the ratio of non-synonymous to synonymous divergence) [123] much less than one, and functionally similar genes are similarly constrained. Moreover, highly expressed genes seem to evolve more slowly. 31.8% of GO categories have significantly lower variance in ω than expected (q-value true-positive test) [124]. Only 11% of GO categories had statistically significantly elevated ω (relative to the median of all genes with GO annotations) at a 5% false-

discovery rate (FDR), suggesting either positive selection or a reduction in the selective constraint. The GO categories showing elevated ω include the biological processes "defence response", "proteolysis", "DNA metabolic process" and "response to biotic stimulus"; the molecular functions "transcription factor activity", "peptidase activity", "receptor binding", "odorant binding", "DNA binding", "receptor activity" and "G-protein-coupled receptor activity"; and the cellular location "extracellular" (Fig.7). When comparing dN (ratio of non-synonymous divergence) across GO categories similar results are obtained, which suggest that in most cases the observed differences in ω among GO categories are driven by amino acid rather than synonymous site substitutions. It is important to notice that the bias in the way biological function are assigned to genes leaves open the possibility that unannotated biological functions may have an important role in evolution. This evidence was backed up by observing that genes with characterized mutant alleles in flybase.org database [5] evolve significantly more slowly than other genes (median ω with alleles =0.0525 and ω without alleles =0.0701; MWU, $P<1x10^{-16}$) [25].

I.5.3. Categories Enriched with Genes Showing Signatures of Adaptive Evolution

There is considerable variation among different protein-coding genes in rates of evolution and patterns of positive selection. Broad functional classes do not seem to share patterns of positive selection, only a few GO categories show excesses of positive selection, although there are some categories who are enriched with genes that show signatures of adaptive evolution such as categories of genes who present interactions with the environment, in sex and reproduction, suggesting that adaptation to changing environments, and sexual selection, shape the evolution of some protein-coding genes [25].

Drosophila species have complex olfactory and gustatory systems used mainly to identify food sources, hazards and mates, which depend on odorant-binding proteins, and olfactory (Ors) and gustatory receptors (Grs). The *D. melanogaster* genome has approximately 60 Ors, 60 Grs, and 50 odorant-binding protein genes. Although there is conservation of gene number across the 12 species, there is evidence that some Or and Gr genes experience positive selection [125-127]. *D. sechellia* and *D. erecta*, two independently evolved specialists, are losing Gr genes approximately five times faster than generalist species [125, 128]. We can easily explain this with the knowledge that generalists are expected to encounter the most diverse set of tastants and because of this they have maintained the greatest diversity of Grs. Indeed, there is some evidence that the odorant-binding protein genes evolve significantly faster in specialists compared to generalists [126]. The observed difference between specialist and generalist ω for Or/Gr genes (0.0292) is significantly greater than the difference for genes across the whole genome (0.0091;

MWU, P=0.0052) [125], suggesting a change in selective regime. The observation that elevated ω affect groups of Or and Gr genes that respond to specific chemical ligands and are expressed during specific life stages, suggests that rapid evolution at Or/Gr loci in specialists is related to the ecological shifts these species have sustained [125].

Also interesting is the case of some detoxification/metabolism genes. The larval food sources for *Drosophila* species contain a significant amount of toxic compounds, consequently *Drosophila* genomes encode a wide variety of detoxification proteins, which include members of the *cytochrome P450* (*P450*) and *glutathione S-transferase* (*GST*) multigene families, all of which are known to have critical roles in resistance to insecticides [129-131]. Among the *P450s*, the five enzymes associated with insecticide resistance are highly dynamic across the phylogeny, with 24 duplication events and 4 loss events since the last common ancestor of the genus, which is in striking contrast to genes with known developmental roles, eight of which are present as a single copy in all 12 species [25]. Specialists seem to lose detoxification genes at a faster rate than generalists (*D. sechellia* has lost the most *P450* genes). And although metabolic enzymes, in general, are highly constrained (median ω=0.045 for enzymes, 0.066 for non-enzymes; MWU, $P=5.7 \times 10^{-24}$), enzymes involved in xenobiotic metabolism evolve significantly faster than other enzymes (median ω=0.05 for the xenobiotic group versus ω=0.045 overall, two-tailed permutation test, P=0.0110) [25].

Drosophila, like all insects, possesses an innate immune system, with many components analogous to the innate immune pathways of mammals although it lacks an antibody-mediated adaptive immune system [132]. Immune system genes evolve rapidly, as a response to selection pressures from pathogens and parasites [133-135]. In the *Drosophila* genus, immune system genes evolve more rapidly than non-immune genes. 29% of receptor genes involved in phagocytosis seem to evolve under positive selection, suggesting that the molecular interactions between *Drosophila* receptors genes and pathogen antigens are causing coevolution and adaptation in these genes. Only 5% of effector genes show evidence of adaptive evolution, compared to 10% of genes genome-wide. It seems, though, that effector genes evolve by rapid duplication and deletion. 49% of genes genome-wide, 63% of genes involved in pathogen recognition and 81% of genes implicated in immune-related signal transduction can be found as single copy orthologues in all 12 species, compared to only 40% of effector genes ($x^2=41.13$, $P=2.53 \times 10^{-8}$). Much of the *Drosophila* immune system genes seems to be evolving rapidly, although the mode of evolution varies across functional classes [25].

Genes encoding sex and reproduction-related proteins are another example of genes subject to a wide array of selective forces, including sexual conflict, sperm competition, cryptic female choice, and all these selective forces are believed to have led to a rapid evolution in these genes [136-138]. These authors [25] analyzed 2,505 sex and reproduction-related genes in the *melanogaster* group, and the data gathered shows that male sex and reproduction-related genes

evolve more rapidly at the protein level than genes not involved in sex or reproduction or than female sex and reproduction-related genes, which comes as an expected result considering the large number of selective forces verified in male flies. Furthermore, genes involved in spermatogenesis have significantly stronger evidence for positive selection than non-spermatogenesis genes (permutation test, P=0.0053), genes that encode components of seminal fluid have significantly stronger evidence for positive selection than 'non-sex' genes [139]. It seems that these selective forces shape the evolution of genomes.

Genes involved in development processes are expected to be under strong negative selection [140]. There is a common generalization of development conservation among related species, with the early development stages being the most refractory to change. Moreover, studies on positive selection in genes related to development are very rare. However, there are positively selected genes found in all embryonic stages, even during the periods who show the highest constraint. Indeed, even in functional groups of genes evolving under strong negative selection such as development, there are many genes showing evidence of positive selection at the amino acid level [140].

I.5.4. The Evolution of Different Genomic Elements

Several genomic elements are evolving at different rates. Here on we mention some of the most important genomic elements in the *Drosophila* genus and how are they evolving.

Gene families expand or contract at a rate of 0.0012 gains and losses per gene per million years, or one fixed gene gain/loss across the genome every 60,000 yr [122]. 342 gene families showed significantly elevated rates of gene gain and loss compared to the genomic average, indicating that non-neutral processes may play a part in gene family evolution [25].

Drosophila cis-regulatory DNA sequences are highly constrained as expected [141-142], with mean constraint within modules as 0.643 (95% bootstrap confidence interval =0.621–0.662) and within footprints 0.692 (0.655–0.723), both of which are significantly higher than mean constraint in non-coding DNA overall 0.555 (0.546–0.563) and significantly lower than constraint at non-degenerate coding sites 0.862 (0.856–0.868) and ncRNA genes 0.864 (0.846–0.880). Further evidencing that transcription factor binding sites tend to be found in larger blocks of constraint that cluster to form cis-regulatory modules [143].

Most of the miRNAs are highly conserved within the *Drosophila* genus. For 71 previously described miRNA genes inferred to be present in the common ancestor of these 12 species, mature miRNA sequences are nearly invariant. Inside the mature miRNA, unpaired sites evolve more slowly than paired sites, whereas the opposite is true for the sequence complementary to the mature miRNA. However, it's interesting to notice that outside the mature miRNA and its

complementary sequence, loops had the highest rate of evolution, followed by unpaired sites, with paired sites having the lowest rate of evolution. A large fraction of unpaired bulges or internal loops in the mature miRNA seem to be conserved. It seems that miRNA genes are considerably more conserved than protein-coding genes. The few substitutions that occur have compensatory changes that maintain the average estimated free energy of the folding structures across the phylogeny [25].

In the mitochondrial genome, functional elements in mtDNA (mitochondrial DNA) are strongly conserved. As expected, tRNAs are relatively more conserved than the mtDNA overall (average pairwise nucleotide distance=0.055 substitutions per site for tRNAs versus 0.125 substitutions per site overall). There is a deficit of substitutions in the stem regions of the stem-loop structure of tRNAs, explained in the light of a strong selective pressure to maintain RNA secondary structure [25].

I.5.5. Positive Selection

Previous works have suggested that a substantial fraction of non-synonymous substitutions in *Drosophila* were fixed through positive selection [144-148]. Estimated 33.1% of single-copy orthologues in the *melanogaster* group have experienced positive selection on at least a subset of codons (q-value true-positive tests) [124]. Thus, several lines of evidence, based on different methodologies, suggest that patterns of amino acid fixation in *Drosophila* genomes have been shaped extensively by positive selection. Among genes inferred to be evolving by positive selection at a 10% FDR, 63.7% (q-value true-positive tests [124]) show evidence for spatial clustering of positively selected codons. In addition to being more constrained, codons in protein domains are less likely to be targets of positive selection [25].

D. melanogaster-specific positive selection has been detected in the past. This could be related to the 'out of Africa' population expansion that occurred approximately 10 000 years ago [149]. *D. melanogaster* expanded from sub-Saharan Africa and colonized other regions of the world and during this expansion it may have come into contact with new pathogens to which the species had to adapt.

Furthermore, many investigators believe that it is very likely that the suggested 10% figure of genes being under positive selection [120] is an underestimation [121].

I.6. Development Studies in *Drosophila*

Several studies using different *Drosophila* species and strains have provided many insights into developmental biology, revealing many genes in the *Drosophila* genome associated with overall developmental processes. Some of these studies are highlighted below, as their contributions to the state of art is worthy of mentioning. First, one important note about development time in the *Drosophila* subgenus is that it varies in different species. Cooley et al. (2012) [95] studied development time in *D. americana* and *D. novamexicana* comparing with identically aged *D. melanogaster* as a control, initiating data collection when the first pupae in the sample reached pupa stage P8 and continuing through adult stage A2. *D. americana* reached the A2 developmental stage 208.5 ± 5.7 hours after the start of the experiment and *D. novamexicana* took a similar amount of time 212.1 ± 3.3 hours. These results were consistent with previously reported qualitative observations of the closely related species *D. virilis* [38]. These developmental times are more than twice that in *D. melanogaster* (98.9 ± 1.7 hours between the start of experiment and occurrence of stage A2). Two sample, two-sided *t*-tests showed that the time taken to reach stage A2 is significantly different between *D. melanogaster* and *D. americana* ($P < 0.0001$) and between *D. melanogaster* and *D. novamexicana* ($P < 0.0001$), but not between *D. americana* and *D. novamexicana* ($P = 0.287$) [95].

Reis at al. (2014) [1] made a F2 association study for developmental time (DT), chill-coma recovery time (CRT), abdominal size (AS) and lifespan (LS) using two *D. americana* strains (H5 and W11), one of which (W11) was a subject of studies in our project. They found significant correlations between DT and LS, that could be explained by variation in genes belonging to both insulin and ecdysone signaling pathways [1]. In their study they analyzed general GO terms selecting for developmental time - GO: 0007476 (imaginal disc-derived wing morphogenesis) and for lifespan - GO: 0008340 (determination of adult lifespan). For the GO term associated with developmental time, 307 putative candidate genes were identified in *D.americana*. 139 genes were identified for the GO term associated with lifespan. In their H5xW11 F2 association study they identified nine molecular markers showing significant associations for developmental time, after applying the sequential Bonferroni correction for multiple comparisons. The strongest signal came from Muller's element C with 8 molecular markers showing significant associations, the other one was identified in Muller's element D. Muller's element C has 64 out of 307 (20.8%) putative candidate genes for the GO term used in this study. This Muller's element is not enhanced by candidate genes, as the relative number (20.8%) is the expected value, considering there are 5 major chromosomes in *D. americana*. However, there seems to be a connection between Muller's element C genes and development time differences [1]. This study also found a significant correlation between DT and LS (DT vs LS, Non-parametric Spearman's correlation =-0.447;

P<0.001), which could indicate that some candidate genes that explain developmental time differences account for variation in other characteristics too, thus pleiotropic genes seem to be involved in the phenotypic differences. Five genes (*EcR*, *ilp2*, *S6k*, *InR*, and *Nf1*) were identified as putative candidates to explain variation in more than one trait. *ilp2*, *S6k*, and *InR* are members of the insulin-signaling pathway that along with *Nf1* have been implicated in the determination of body size [150-152] and adult lifespan [153-156]. However, only *EcR*, located on Muller's element C, is a putative candidate for variation in both DT and LS. *EcR* has been shown to have a role in development during metamorphosis [157] and it is involved in the regulation of longevity [158]. This goes in accordance with the already reported impact of ecdysone signaling pathway in development [1]. It seems very likely that, in *D. americana*, the variation in DT is due to the involvement of the pleiotropic gene *EcR*, but also due to multiple other genes located across the entire genome [1]. None of these five genes show signs of positive selection according to the bpositive online platform [110].

Costas et al. (2003) [159] identified and characterized a repetitive motif in the genome of *D. melanogaster* that shows a strong association with developmental genes in the genus. The *wingless (wg)* gene located in the chromosome 2L of *D.melanogaster* is a member of the *Wnt* gene family that encode for secreted glycoproteins, which act as key intercellular signaling molecules during animal development [160]. While searching the *D. melanogaster wg* intron sequences for putative regulatory regions, using an evolutionary comparative approach, the authors identified a 27 bp long motif that is overrepresented within the *D. melanogaster* genome and that is strongly associated with genes involved in development, signal transduction, and transcriptional regulation. This motif could prove useful in searching for new genes involved in *Drosophila* development as several components of main signaling pathways are associated with it, such as: *Delta* (*Dl*) and *Serrate* (*Ser*) (ligands), *Notch* (*N*) (receptor) and *Su(H)* (nuclear transducer) of the *N* signaling pathway; *wg* (ligand) and *frizzled3* (*fz3*) (receptor) of the *Wnt* signaling pathway; *Epidermal growth factor receptor* (*Egfr*) (receptor) and *vein* (*vn*) (ligand) of the *Egfr* signaling pathway; *hedgehog* (*hh*) (ligand) of the *hh* signaling pathway; or *decapentaplegic (dpp)* (ligand) of the *TGF*-β receptor signaling pathway. Thus, it's likely that this motif defines an important regulatory network, linking together several fundamental genes active during *Drosophila* development [159]. Several characteristics of the motif, such as its trend to form clusters within and around genes or its biased location in regard to the transcription units, might be a consequence of its association with regulatory regions of genes associated with signal transduction pathways, and transcription factors involved in several developmental processes [159]. Furthermore, the detection of this motif in other *Drosophila* species from the *Drosophila* subgenus shows that this motif arose within the genome before the radiation of the genus [161-162] which might suggest an evolutionary conservation linked with its probable vital importance during *Drosophila* development [159].

Yan et al. (2015) [163] verified that the loss of *GPAT4* in *Drosophila* led to a severe delay of development and slow growth, resulting in small-sized animals during the larval stages, however resulting ultimately in normal sized adult flies. This striking evidence, suggests that there are robust body size adjustments in the *GPAT4* mutant likely achieved by changes in *ecdysone* and *insulin* signaling, and also in compromised food intake. It seems highly probable that a strategy evolved in animals to reach final body size when challenged by genetic alterations. Accordingly, a perturbed wing disc in *Drosophila* would slow down the growth of other discs and as a result the whole body, presumably allowing enough time for the damaged disc to repair [164-166]. During this repair process, *ecdysone* signaling is considered to be the limiting factor that ensures a normal final body size [165]. Previous studies show that perturbed wing discs secrete *dilp8*, an insulin-like peptide that inhibits ecdysone biosynthesis and slows down the growth of other tissues [167-168]. Likewise, loss of *Ceng1* results in normal-sized adult flies with a moderate delay of the second instar larval stage in *Drosophila* [169]. The mutants show reduced expression of *ecdysone*, which presumably explains the extension of developmental time [169]. *Ecdysone* has repeatedly been considered a pivotal hormone that controls the development timing in *Drosophila* [170-172]. Besides the above-mentioned study by Reis et al. (2014) [1], other studies have shown the association of *ecdysone* with developmental time, reduced *ecdysone* signaling has been reported to increase the final body size in *Drosophila* through accelerating the growth rate or extending the duration of development [170-172]. Body growth is tightly associated with nutrient status and nutrient demand in all organisms [173-175], and it is known that for a similar nutrient demand, a high level of nutrients would promote organismal growth whereas a low level of nutrients would limit organismal growth [173-175]. It is also known that at a similar nutrient level, the nutrient demand is connected to the physiological and genetic conditions of the organism. The *ecdysone* signal and the *insulin* signal are critical signals that control growth duration and growth rate, respectively [173-175].

This study further evidences the associations between developmental time and lifespan, as well as the association between *insulin* and *ecdysone* signaling pathways and developmental time. It is likely that every single one of genes associated with *ecdysone* and *insulin* signaling pathways might have a different and essential role in development processes and consequently in developmental time.

II. Material and Methods

II.1. Bioinformatic Search of Candidate Genes and the bpositive Database

We performed a search using flybase.org database [5] for *Drosophila* genes with molecular functions and biological processes related with development, overall development, development processes and developmental time. Using the tool "biological process" in flybase.org we were able to find several GO (Gene Ontology) terms related with development processes and make several different basic analyzes of the chromosomal location of genes in each GO term list.

To analyze if genes present any sign of positive selection, we used the online platform bpositive (http://bpositive.i3s.up.pt/) [110] dataset BP2016000001. This platform uses data on the *D. melanogaster* genome annotation (release 6.12) and data on the genome assemblies of 11 closely related *Drosophila* species, namely: *D. simulans* (release 1.4), *D. sechellia* (release 1.3), *D. erecta* (release 1.3), *D. yakuba* (release 1.3), *D. suzukii* (release 1), *D. biarmipes* (release 1), *D. takahashii* (release 1), *D. eugracilis* (release 1), *D. ficusphila* (release 1), *D. elegans* (release 1) and *D. rhopaloa* (release 1). The annotation from all *D. melanogaster* coding sequences (CDS) was transferred to the other 11 species, using Compart and Splign, as implemented in BDBM (BLAST DataBase Manager). Only fully annotated CDS, with no in-frame stop codons and with maximum 10% length variation relative to the *D. melanogaster* reference sequence were kept. Thus, it should be noted that rapidly evolving genes and genes encoding proteins with a variable number of repeats are likely not annotated this way. However, it is well established that it is difficult to obtain a reliable alignment for genes with such features. Moreover, it is well known that the use of highly divergent sequences can lead to an underestimation of the rate of synonymous (nucleotide change resulting in no alteration of the produced amino acid) divergence, which in turn leads to an inflated value for the ratio of the rate of non-synonymous (nucleotide change resulting in alteration of the produced amino acid) divergence/rate of synonymous divergence, leading to the identification of most likely false positively selected amino acid sites. The database uses ClustalW2, MrBayes, and codeml, as implemented in ADOPS [176], to attempt to identify positively selected amino acid positions in all datasets with four or more sequences. The resulting ADOPS project folders are provided, and all details of the performed analyses can be inspected, moreover, researchers can do additional analyses on the data, for instances, assessing the impact of the inclusion of a given sequence, or the impact of using a different alignment algorithm [110].

BP2016000001 was the used dataset in our project. Results for a specific gene are available for all the different transcripts available, and are presented as "analyzed", for genes who were able to be analyzed but don't show signs of positive selection; "not analyzed", for genes who were not able to be analyzed due to the way the annotation was done; and "positively selected", for genes that not only are "analyzed" but also show signs of positive selection. Furthermore, we looked at the sequences of "positively selected" genes in order to verify whether or not we believe in the prediction of the software. Although it is a subjective analysis, some transcripts could have resulted in "positively selected" as a result of poor alignment or the use of a small number of sequences.

II.2. *Drosophila* Strains and Husbandry

In this project we used 3 different strains of *Drosophila*:

- *D. americana* strain W11
- *D. americana* strain SF12
- *D. novamexicana* strain 15010-1031.00 (shortly mentioned as NOVA00)

Stocks from the flies were kept in laboratory conditions. During all experiments, all flies were raised under similar conditions, on standard cornmeal food with yeast at 25ºC.

II.3. Experimental Setup

We collected samples from the 3 strains with different development times and at different stages/days of life to later perform RT-PCR and study the expression of the candidate genes in the different samples. All samples are made entirely of male flies. A total of 12 different conditions were set and later tested: W11 flies with 15 days of development at 4 days and 8 days of life, W11 flies with 16 days of development at 4 days and 8 days of life, SF12 flies with 15 days of development at 4 days and 8 days of life, SF12 flies with 16 days of development at 4 days and 8 days of life, NOVA00 flies with 16 days of development at 4 days and 8 days of life and NOVA00 flies with 17 days of development at 4 days and 8 days of life. The different conditions used for NOVA00 are due to the difficulty in obtaining flies with 14 and 15 days of development. This observation agrees with the suggestion of Cooley et al. (2012) [95] showing that *D. novamexicana* has a longer development time than the sister species *D. americana* when comparing the time taken to develop from Pupal stage P8 to Adult stage A2.

In order to obtain flies with the specific day of development, firstly, flies from stocks were collected on the day of their birth. They were then transferred to another tube for 7 days (no crosses between species or strains were performed, all the newborn flies were transferred to a new tube) and after this time transferred to another new tube to lay eggs for 24 hours (the objective was to make sure that all the progeny had 7 days, so the age of progeny wouldn't have an effect on the outcome of the experiment). The larva deposited in this tube gave origin to adults after 14-18 days (development time days, from egg to adult). All male flies were collected the day they were born. They were kept in these tubes for 4 and 8 days before being snap frozen using liquid nitrogen and kept at -80ºC in sets of three. Sets of three males per tube were used for each condition to guarantee an RNA concentration high enough for further cDNA synthesis. The minimum number of biological replicates is 3, however, in most cases, more sets were used.

We selected 4 and 8 days of life to snap froze flies since gene expression usually varies with age and it is unclear what the best time to assay differences is. We did not select newly born flies or flies with 1 day of life because the gene expression in the early stages of adult life can be remarkably variable (the difference between few hours of life can impact the expression producing significant differences). By using 4 and 8 days we can also verify whether or not there is an increase or decrease in the expression of the genes studied during adult life.

II.4. RNA Extraction

Total RNA was isolated from the whole body of the collected males of each sample, each tube containing a total of three males with the specific condition in study. Ambion™ TRIzol™ Reagent (Ref.15596018, Invitrogen, Carlsbad, CA, USA) protocol was optimized in order to improve the quality of the extracted RNA samples, improving the overall RNA integrity and reducing the contamination with DNA. The final protocol is the following: homogenize the samples using 100 µL of TRIzol™. After homogenization add another 900 µL of TRIzol™ reagent to the samples. Incubate the samples 5min at room temperature (RT). Centrifuge 10min at 5000 rpm in 4ºC. Transfer the supernatant to another 1.5 mL tube. Add 200 µL of Chloroform: Isoamyl Alcohol 24:1 (C0549 SIGMA, St. Louis, Missouri, USA). Stir the samples for 3 min at room temperature. Move the samples into 5Prime Phase Lock Gel™ Heavy 2 mL tube (#2302810, Quantabio, Hamburg, Germany). Centrifuge 15 min at 5000 rpm at 4ºC. Transfer the supernatant to another 1.5 mL tube. Add 5 µL of RNase-free Glycogen (5 mg/mL) (Ref.AM9510, Invitrogen, Lithuania) and invert. Add 500 µL of 2-Propanol for Analysis Emsure® (Ref.1.09634.1000, MilliporeSigma, Darmstadt, Germany). Incubate 10 min at RT. Centrifuge 10 min at 5000 rpm at 4ºC. Remove and discard supernatant. Wash the RNA pellet twice with 1 mL of 75% ethanol with

Ambion™ Nuclease-free Water (not DEPC-Treated) (Ref.AM9932, Invitrogen™, USA) by centrifuging for 5 min at 5000 rpm at 4ºC. Remove and discard supernatant. Let the samples dry on air for 10 min in ice, use scientific paper to remove excessive water/reagents. Dissolve the RNA pellet in 21.5 µL of Ambion™ Nuclease-free Water and resuspend the sample. The RNA is kept at -80ºC.

II.5. DNA Digestion

The RNA samples were digested/treated with TURBO DNA-free™ Kit (Ref.AM1907, Invitrogen™, Vilnius, Lithuania) using the following adapted protocol with kit reagents: Add 2.5 µL of 10xTurbo DNase™ buffer and 1.2 µL of Turbo DNase™ to the tube with 21.5 µL, for a total of 25.2 µL. Incubate at 37ºC for 30-45 mins. Add 2.5 µL of DNase™ inactivation reagent and stir samples. Incubate for 5 min at RT with constant movement. Centrifuge at 5000 rpm for 1.5 min. Transfer the RNA to a new tube and store at -80ºC.

II.6. Assaying DNA Contamination

We performed PCR to check for DNA contamination of the samples. The samples used on this PCR were diluted by adding 1.0 µL of RNA to 1.5 µL of H_2O. This is the same concentration used further for cDNA synthesis. The overall mix was made according to the following 1x recipe: 5.95 µL H_2O; 1.2 µL 25mM Magnesium Chloride (Ref. R0971, Thermo Fischer Scientific, USA); 1 µL 10x Taq Buffer with $(NH_4)_2SO_4$ (Ref.B33, Thermo Fischer Scientific, USA); 1µL dNTPs; 0.125 µL PF (primer forward); 0.125 µL PR (primer reverse); 0.1 µL 5 U/µL Taq DNA Polymerase recombinant enzyme (Ref.10342020, Invitrogen™, USA) making a total of 9.5 µL. For each tested sample add 0.5 µL of RNA for a total of 10 µL volume per tube, the remaining 9.5 µL of the mix. The primers used were LCO (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') [177]. Positive controls were performed with genomic DNA from the strains studied (W11, SF12, NOVA00). The PCR was run using the protocol: 96ºC for 2 min, 40 cycles of 95ºC for 30s, 50ºC for 45s, and 72ºC for 2 min, followed by 72ºC for 5 mins, using Biometra TPersomal Thermal Cycler (analytikjena©, Gottingen, Germany). Results were observed using Gibco BRL Horizon 11-14 Horizontal Gel Electrophoresis (Gel Company, San Francisco, CA, USA). The DNA amplifications were observed using GeneFlash BioImaging (SyngeneBioImaging, Cambridge, United Kingdom).

Samples showing evidence for DNA contamination show similar results to the positive controls (a band around the size of the amplified product, in this case, 700 bp band on a 1.5% agarose gel). Samples without DNA contamination show similar results to the negative controls, that is no amplification is expected. Samples with no apparent DNA contamination are subjected to RNA integrity test using Experion™ Automated Electrophoresis System (Bio-Rad, Portugal) protocol. Samples showing evidence for DNA contamination are discarded.

II.7. Assaying RNA Integrity

RNA integrity, quality, and concentration were assessed with the use of Experion™ Automated Electrophoresis System (Bio-Rad, Portugal). Tested samples were diluted by adding 0.5 µL RNA to 1 µL H2O. They were run in Experion™. The samples with "green" RQI (RNA quality indicator) (RQI>7) are used for further cDNA synthesis, while samples who didn't meet this criteria were discarded.

II.8. cDNA Synthesis

We performed cDNA synthesis, of the samples with RQI>7, using reverse transcription of 1.0 µg of RNA of each sample with SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR (Ref.18080400, Invitrogen™, Carlsbad, CA, USA) using random primers, according to the manufacturer's protocol. One reaction where the template was not added (RT+) was performed. Reactions with RNA that was not reverse transcribed (RT-) were performed for each RNA sample to confirm the absence of genomic DNA contamination.

II.9. RT-PCR

Real-time PCR (RT-PCR) was performed for all cDNA samples using specific primers for the candidate genes studied and the reference gene *RPL32*. Moreover, RT- samples were added to the PCR plate. For each of the samples, we performed two technical replicates. The housekeeping gene *RPL32*, also known as *Ribosomal Protein L32* or *rp49*, is commonly used in *D. melanogaster* as a reference gene for RT-PCR studies [178]. RT-PCR was performed on iQ™ 96-Well PCR Plates, (Ref.2239441, Bio-Rad, Portugal) with a total 10 µL in each well. This consisted of 6.5 µL iTaq™ Universal SYBR® Green Supermix (Ref. 1725122, Bio-Rad, Portugal), 3.75 µL H_2O, 0.125 µL Primer Forward, 0.125 µL Primer Reverse and 1 µL of diluted

cDNA samples. The dilutions were made according to the used primers (*Act79B* - 1:10; all other genes – 1:5).

RT-PCR was performed on a CFX Connect™ Real-Time PCR Detection System (Ref. 1855200, Bio-Rad, Portugal) with the following program: 3 min at 95°C; 40 cycles of 30s at 94°C, 30s at 56°C and 30s at 72°C followed by a standard melt curve and 3 min at 95°C. The cDNA levels were normalized to *RPL32* levels in the same samples. Relative expression values were determined by the $2^{-\Delta CT}$ method [179] and the two-sample two-sided Student's t-test was applied to the data to address if there are statistically significant differences in relative gene expression between different conditions/samples.

II.10. Primers Used for RT-PCR

The primers used for real-time PCR were obtained using the software OLIGO version 1.4 (National Biosciences, Inc.) based on the *D. americana* and *D. novamexicana* genome sequence. The expression of these genes was studied in these two non-organism models, for which there are no premade primers for RT-PCR for any gene in the genome. There are sequence differences in the genomes of the different strains studied, and in the genomes of different individuals from the same strain, and these sequences cannot be used. Moreover, primers were selected if they followed the following criteria: have amplicon length between 50-210 bp; a length from 19-23 nucleotides; avoid primer self or cross annealing stretches > 4 bp; dimer strengths with ΔG value more positive than -9 kcal/mol; software predicted hairpin Tm (melting temperature) at 5ºC below annealing temperature; GC content from 35-65%; and close Tm between reverse and forward primers.

The primers were confirmed as having a unique pairing sequence using BLAST search on flybase.org and NCBI blast search against the available *D. americana* and *D. novamexicana* sequences. PCR was performed on genomic DNA first, to confirm the presence of only one amplified product, using gel electrophoresis. The primer pairs that give rise to only one amplification product, were tested on cDNA to confirm their amplification of only one fragment in cDNA. We tested the primers amplification on all the different strains cDNA. The efficiency of the primers was measured using serial dilutions of cDNA (1:1; 1:5; 1:25 and 1:125) from a mix of samples from the 3 different studied strains. Considering manufacturers ThermoFisher Scientific recommendations and the knowledge on RT-PCR in *Drosophila* species, we worked with primers with efficiency level between 90-100%. We developed several different primers for the candidate genes and for other genes, however not all the designed primers had efficiency levels between 90-100% (Supplementary Table 1). We obtained primers with the required efficiency for RT-PCR amplification for six candidate genes (Table 2).

Table 2- Sequence of primers forward and reverse for the 6 candidate genes and the reference gene (*RPL32*). Primer sequence (5'→3')

Gene	Primer Forward	Primer Reverse
Act79B	CTACTCGTTCACCACCAC;	TACCGCCAGACAGCACAT
hpo	AAACTGGGCGAGGGTAGC;	CCACATACGGGGAATCACA
yki	ACCAGCCATCACACAGTC;	TAGCGAAGTCCACATTATTAT
Sin3A	GTCTCGTTTCCTACCTGGGC;	CGCTCGTCCTCCGTTCTGTA
sm	TTCAACCACCGAGCAAAATCT;	TCAAGGGAAAGAGACGCACAG
CG3209	CGAAGACGATGTTACCTC;	CCAATAAAGCACACTAAGACG
RPL32	ACAACAGAGTGCGTCGTC;	ATCTCCTTGCGTTTCTTC

II.11. Electrophoresis Gels

All electrophoresis gels were 1.5% agarose gels using TAE (Tris-Acetate-EDTA) buffer. To make a complete gel we add 2.4 g agarose to 120 mL deionized water, warmed in the microwave for around 1-2 min. After which we added another 24 mL of deionized water, 16 mL of Ambion™ 10x TAE (Ref.AM9869, Invitrogen™, Portugal) and 3 µL of GreenSafe™ Premium stain (Nzytech, Lisboa, Portugal).

III. Results

III.1. Selection of Candidate Genes

Several approaches can be taken in order to identify the molecular basis of phenotypic variation, such as focusing on associations between genetic variation within genes of interest and phenotypes. This can be approached by both genetically modifying the candidate genes, as well as examining the natural variation of the candidate genes, the latter being the line of work we followed. Most of candidate gene approaches rely on the previous knowledge of the genes biological functional and putative impact on the trait in question.

Indeed, in our work, the goal was to find variation in candidate genes that could be associated with differences in developmental time in *Drosophila.* The reported association between genes and *Drosophila* developmental time variation is unsatisfying, it is questionable what are the best criteria to follow and what categories of genes likely have a better link with development time in *Drosophila.* Our search relied on 3 main criteria. The first one was to find genes associated with development time, namely, in GO (Gene Ontology) terms lists from flybase.org [5], related with development, as well as bibliographic evidence of the connection of the genes with developmental processes reported by previous studies, using databases such as NCBI and Elsevier. The second criterion was to focus on finding genes located on Muller's element C, which corresponds to chromosome 2R in *D.melanogaster* and corresponds to chromosome 5 in *D. virilis*. As mentioned before, Reis et al. (2014) [1], performed an F2 association study for development time (DT), chill-coma recovery time (CRT), abdominal size (AS) and lifespan (LS) and reported nine molecular markers showing significant associations with development time, eight of those being present on Muller's element C. Therefore, there seems to be a connection between Muller's element C genes and development time. The last criterion was to find genes showing signs of positive selection. As development genes are under strong negative selection [140], therefore, the ones who show signs of positive selection are interesting candidates to explain the phenotypic differences in development time. We used bpositive online platform (http://bpositive.i3s.up.pt/) [110] to search for signs of positive selection in the candidate genes.

We analyzed 71 lists of GO terms related with development and performed analyses of chromosomal locations of this genes in order to detect if any list was enhanced with genes belonging to Muller's element C (Supplementary Table 2). Each of the GO term lists was analyzed by number of matches, number of matches/genes associated with a chromosome (take into consideration that some GO list terms are filled with homolog genes for different *Drosophila* species (not *D. melanogaster*) and/or genes present in small scaffold not assigned to any Muller's

element), and absolute and relative number of genes present in each one of the six chromosomes (3R, 2R, 3L, 2L, X and 4). Furthermore, we present the ChiTEST p values for all the GO Term lists, and the ChiTEST p values when not considering chromosome 4. For the expected number of genes, we based our work on the amount of known coding genes in each Muller's element, presented by the *Drosophila* 12 Genomes Consortium on "Evolutions of genes and genomes in the *Drosophila* phylogeny" [25].

Our analysis suggests that no list containing a significant amount of genes (more than 200 genes) is enhanced with more genes than what was expected for the chromosome 2R. However, we observed significantly different results than expected for six of the smaller (less than 200 genes) GO term lists "Brain development"; "Muscle cell development"; "Response to ecdysone"; "Response to insulin"; "Ectoderm development" and "Muscle fiber development". When we don't consider chromosome 4 (as this chromosome represents only 1% of the known coding genes in the *D. melanogaster* genome) the ChiTEST for the GO term "Brain development" no longer shows a statistically significant difference. The GO terms "response to insulin" and "response to ecdysone" show significant differences, however, this is not a result of an enhancement of genes on chromosome 2R. Moreover, from the six lists with significant ChiTEST p values, only "Muscle fiber development" shows enhanced number of genes on chromosome 2R, however, the list consists only in 12 genes, therefore it seems unlikely that the differences have any significant biological meaning. These analyses suggest that no GO term list related with developmental processes is overrepresented in the chromosome 2R, and only six small GO term lists have significantly different observed distributions of genes per chromosome than expected.

A subset of GO terms (GO:0009888-"tissue development", GO:0048856-"anatomical structure development", GO:0009790-"embryo development", GO:0048589-"developmental growth", GO:0035220-"wing disc development", GO:0032502-"developmental process", GO:0008340-"determination of adult lifespan", GO:0035075-"response to ecdysone", GO:0032868-"response to insulin") were considered specially interesting, as they consist of lists with bigger number of genes and/or GO terms that might more accurately reflect differences and association with developmental time, and thus were further analyzed. We add the remaining criteria in order to narrow down the search, analyzing these GO term lists by number of total genes present in the list, number of genes located on Muller's element C and from genes on this element: number of "analyzed" genes by the bpositive platform and number of "positively selected" genes on bpositive platform, i.e., genes showing signs of positive selection according to the results in bpositive platform. The results are presented below in Table 3 relative to the *D. melanogaster* matches.

Table 3 - Number of *D. melanogaster* genes in each one of the selected GO Terms, by GO TERM; the number of genes associated with a chromosomal arm; number of genes located on chromosome 2R; number of these 2R genes that are analyzed according to bpositive (b+); number of these analyzed genes in 2R that show signs of positive selection according to bpositive (P+). Lastly, the ratio between genes showing positive selection located on chromosome 2R and total of genes on the GO Term list.

GO TERM	N. of genes	2R	Analyzed	P+	2R & P+/Matches
developmental process	3305	651	495	95	3%
anatomical structure development	3075	622	479	88	3%
tissue development	1122	207	150	37	3%
embryo development	556	115	94	14	3%
wing disc development	401	83	72	22	5%
developmental growth	380	68	57	12	3%
determination of adult lifespan	166	31	26	4	2%
response to ecdysone	52	8	6	2	4%
response to insulin	52	3	3	2	4%

A complete list of the genes belonging in each of the above GO term lists for *D. melanogaster*, present in Muller's element C and showing signs of positive selection is presented below in Tables 4-12.

These were chosen as the main candidates for our work. But as we can observe in Table 3, even though we managed to reduce the number of candidate genes (to 5% or less than the original lists), there are still too many putative candidate genes to study.

Table 4 - Positively selected genes from GO:0032502-"developmental process" on chromosome 2R of *D. melanogaster*

Genes							
Amph	*clu*	*emp*	*Hr3*	*mbl*	*pdm3*	*px*	*SP2353*
CAP	*cora*	***Eps-15***	*Hr51*	*MFS14*	*Pfk*	***retn***	*sqa*
CG3209	*coro*	*fj*	*imd*	***Mmp1***	*Pi3K59F*	*sano*	***stj***
CG5742	*Cp1*	*fus*	*insc*	*NAT1*	*PIG-V*	*scra*	*tamo*
CG6145	*Cpr51A*	*Gapdh1*	*jeb*	*nemy*	*pio*	*Sdc*	*tvv*
CG12484	*cv-2*	*geminin*	*Khc*	*nompA*	***Pip5K59B***	*Sema2a*	*Unc-89*
CG13531	*Cyp301a1*	***GLS***	*kn*	*Not1*	*pk*	*seq*	*ValRS*
CG42663	*Dll*	*Gp150*	*l(2)dtl*	***Nurf-38***	*Pka-R2*	*SERCA*	*VepD*
CG43729	*Drep1*	*grh*	*Lac*	*olf186-F*	*Pkn*	*sha*	*Vmat*
Chi	*Dscam1*	*Galphaq*	*Lcp9*	*Pal1*	*Prosap*	*Shroom*	***yki***
Chn	*egl*	*hbs*	*lola*	*par-1*	*psq*	*Slik*	*Zasp52*
ckn	*Ehbp1*	*hig*	*Lpin*	*Patronin*	*Pu*	***sm***	

Table 5 - Positively selected genes from GO:0048856-"anatomical structure development" on chromosome 2R of *D. melanogaster*

Genes							
CG6145	*clu*	*emp*	*Hr51*	*mbl*	*Pi3K59F*	***retn***	*sqa*
Amph	*cora*	***Eps-15***	*Hsc70-5*	*MFS14*	*PIG-V*	*sano*	*stan*
CAP	*coro*	*fj*	*Ih*	***Mmp1***	*pio*	*Sdc*	***stj***
CG13531	*Cp1*	*Galphaq*	*imd*	*nemy*	***PIP5K59B***	*Sema2a*	*tamo*
CG3209	*Cpr51A*	*Gapdh1*	*jeb*	*nompA*	*pk*	*seq*	*Tor*
CG42663	*Cyp301a1*	*geminin*	*Khc*	*Not1*	*Pka-R2*	*SERCA*	*ttv*
CG5742	*Dll*	***GLS***	*kn*	*Pal1*	*Pkn*	*sha*	*ValRS*
CG6191	*Drep1*	*grh*	*l(2)dtl*	*par-1*	*Prosap*	*Shroom*	*VepD*
Chi	*Dscam1*	*hbs*	*Lac*	*Patronin*	*psq*	*Slik*	*Vmat*
chn	*egl*	*hig*	*lola*	*pdm3*	*Pu*	***sm***	***yki***
ckn	*Ehbp1*	*Hr3*	*Lpin*	*Pfk*	*px*	*SP2353*	*Zasp52*

Table 6 - Positively selected genes from GO:0009888-"tissue development" on chromosome 2R of *D. melanogaster*

Genes						
CG6191	*fj*	*insc*	*mim*	*Pde8*	*px*	***Sin3A***
Chi	*geminin*	*jeb*	***Mmp1***	*pio*	*sano*	*stan*
Cora	***GLS***	*Khc*	*nemy*	*pk*	*seq*	*ths*
cv-2	*grh*	*kn*	*Not1*	*Pkn*	*SERCA*	*tor*
eEF5	*hbs*	*Lac*	*Pal1*	*psq*		
ena	*Hr51*	*MFS14*	*par-1*			

Table 7 - Positively selected genes from GO:0035220-"wing disc development" on chromosome 2R of *D. melanogaster*

Genes						
Chi	*eeF5*	*Hr51*	*Lpin*	*nemy*	*pio*	*psq*
Cora	*fj*	*kn*	*mbl*	*Pal1*	*pk*	*px*
cv-2	***GLS***	*lola*	*MFS14*	*par-1*	*Pkn*	*sha*
Dll						

Table 8 - Positively selected genes from GO:0009790-"embryo development" on chromosome 2R of *D. melanogaster*

Genes						
Chi	*egl*	*geminin*	*Khc*	*Lac*	*psq*	***retn***
cora	*eve*	*jeb*	*kn*	*par-1*	*Pu*	*stan*

Table 9 - Positively selected genes from GO:0048589-"developmental growth" on chromosome 2R of *D. melanogaster*

Genes					
CG13531	***Eps-15***	***Mmp1***	***PIP5K59B***	*Sdc*	***stj***
Dscam1	*fj*	*par-1*	*Prosap*	*stan*	*Tor*

Table 10 - Positively selected genes from GO:0008340-"determination of adult lifespan" on chromosome 2R of *D. melanogaster*

Genes			
CG42663	***Sin3A***	***Ih***	***sm***

Table 11 - Positively selected genes from GO:0035075-"response to ecdysone" on chromosome 2R of *D. melanogaster*

Genes	
CG3209	***Nurf-38***

Table 12 - Positively selected genes from GO:0032868-"response to insulin" on chromosome 2R of *D. melanogaster*

Genes	
CG3209	***hppy***

We choose a subset of genes from all candidate genes present on the GO term lists on Tables 4-12, by enriching the analyzes of GO terms with bibliographical searches. Highlighted on Tables 4-12, are the genes for which we designed primers, however, as mentioned before in Material and Methods, not all the genes were studied as a consequence of primers efficiency test results.

In the work of Reis et al. (2014) [1], there is a suggestion of a correlation between development time and determination of adult lifespan. Therefore, we considered this second term. When we performed analyses of this GO list, only 4 genes match our criteria (*Sin3A*, *sm*, *CG42663*, and *Ih*). In this work, we studied *Sin3A* and *sm*, as both of them show previous association with development according to other authors and are present in other GO term lists related with development. *CG42663* is only associated with GO term "determination of adult lifespan" therefore we opted to not study its expression, as there is little previous knowledge on this gene. Moreover, *Ih* was one of the candidate genes for which we couldn't obtain efficient primers for amplification in the 3 strains on RT-PCR.

After narrowing down the search and designing primers we managed to study the expression of four genes (*Sin3A*, *sm*, *CG3209*, and *yki*) that followed strictly the criteria set in this work. Besides, the main three criteria set, interestingly, *Sin3A* and *sm* are present under GO term: "determination of adult lifespan". *CG3209* is present in the GO term: "response to ecdysone" and GO term: "response to insulin". *CG3209* is a reported gene involved in the insulin and ecdysone signaling pathways, and considering the work by Reis et al. (2014)[1] these pathways are related with development. Lastly, *yki* who also belongs to the *hippo* signaling pathway, and according to Wittkorn et al. (2015) [180], "The evolutionarily conserved *hippo* signaling pathway is known to regulate cell proliferation and maintain tissue homeostasis during development", *yki* is one of the

essential genes present in this pathway, that according to bpositive, shows signs of positive selection.

Taken this into account we also studied *hpo*, although it is not a gene showing signs of positive selection according to the bpositive platform, it is, however, present on Muller's element C, and it's the main gene on the *hippo* signaling pathway.

We also studied the expression of *Act79B.* Although not located on Muller's element C, and not showing signs of positive selection, it has been included as Reis et al. (2015)[181] study suggest that its expression levels can be used as a molecular marker of biological age, and ontogenetic delay. These processes, as well as lifespan, have been associated with development time [1]. Considering that *Act79B* might act as a marker of molecular biological age, the study of its expression in flies showing different development time might provide some insights into whether or not flies that take more time to develop are biologically younger than flies who take less time to develop, as it's the case with diapausing *D. americana* females [181].

The biographical data supporting the decision to study each one of the six genes is presented in detail below.

III.2. The Candidate Genes

III.2.1. Sin3A

The gene *Sin3A* is located in Muller's element C and shows signs of positive selection on 5 of its 10 know *D. melanogaster* transcripts according to bpositive database BP2016000001 [110]. These are the five transcripts with lengthiest coding sequences (around 6189 bp compared to around 5300 bp for the transcripts who don't show signs of positive selection). The known *D. virilis* homolog is named *GJ20103*. *Sin3A* is associated with GO terms related to development such as "larval somatic muscle development"; "muscle organ development"; "neuron development", "negative regulation of *Wnt* signaling pathway"; "wing disc development"; "tissue development" among other GO terms. It also appears in GO term: "determination of adult lifespan" being one of only 4 out of 166 genes in this list that is present in Muller's element C and shows signs of positive selection. *Sin3A* is part of the *SIN3* complex, one of two major class I containing histone deacetylases (HDAC) complexes present in cells [182]. *Sin3A* is considered a transcriptional corepressor essential in metazoans, more precisely a scaffolding protein that complexes with HDAC where it interacts with corepressors to control transcriptional silencing of genes [183].

In cell culture experiments, the loss of the *SIN3* complex leads to defects in cell proliferation and affects the normal cell growth [184]. Mutations in this complex result in lethality in mouse model systems [185-189]. Null mutations of *Sin3A* in *Drosophila* results in embryonic lethality with only a few animals surviving to the first larval instar stage [188-189]. In *Drosophila* tissue culture cells, reduction of *Sin3A* protein expression by RNAi (RNA interference) resulted in a G2 phase delay in cell cycle progression [190]. A comparison of gene expression profiles from wild-type and RNAi-induced *Sin3A* knockdown cells revealed differences in expression of genes encoding proteins that control multiple cellular processes, including cell cycle progression, transcription, mitochondrial activity and signal transduction [191]. Expression of two genes critical for the G2/M transition of the cell cycle, *STG (String*) and *CycB (cyclin B*), was reduced in the *Sin3A* knockdown tissue culture [187]. The *SIN3* complex is also necessary for post-embryonic development [192], loss of *Sin3A* from wing imaginal disc cells resulted in a number of observable phenotypes, including smaller imaginal discs and smaller, curly adult wings. To summarize, *Sin3A* is an essential gene required for both embryonic and larval development [188-189, 192]

III.2.2. sm (smooth)

The gene *sm*, also known as *smooth* is located in Muller's element C and shows signs of positive selection on 3 of its 23 known *D. melanogaster* transcripts according to bpositive database BP2016000001 [110]. These are the three transcripts with lengthiest coding sequences (around 1600 bp compared to around 780-1500 bp for the transcripts who don't show signs of positive selection). The known *D. virilis* homolog is named *GJ21415*. *sm* is present in GO terms related to development such as "axon guidance"; "mRNA processing"; "anatomical structure"; "developmental process" among others. It also appears in GO term: "determination of adult lifespan" being one of only 4 out of 169 genes in this list that is present in Muller's element C and shows signs of positive selection. s*mooth* (*sm*) is the homolog of the human heterogeneous nuclear ribonucleoprotein L (*hnRNPL*), and is involved in mRNA synthesis and maturation. *sm* is primarily expressed in chemosensory neurons and homozygous *sm* mutants show defects in axonal arborization of chemosensory neurons and in feeding behavior. Feeding behavior is crucial for development. These defects may be related to their early death after eclosion [193-194].

III.2.3. CG3209 (GPAT4)

The gene *CG3209*, also known as *GPAT4* is located in Muller's element C and shows signs of positive selection on 2 of its 4 known *D. melanogaster* transcripts according to bpositive database BP2016000001 [110]. The four transcripts show variation in coding sequence length, with the lengthiest one (1985 bp) and the shortest one (1660 bp) showing signs of positive selection. The known *D. virilis* homolog is named *GJ20434*. *CG3209* is present in GO terms related to development such as "developmental process"; "instar larval development"; "anatomical structure" among others. It also appears in GO term: "ecdysone receptor-mediated signaling pathway" and "response to insulin", two important pathways that are related with development time differences, as previously suggested by Reis et al. (2014)[1], Yan et al. (2015)[163], Colombani et al. (2005)[170], Layalle et al. (2008)[171], Tennessen et al. (2011)[172]. The mutation of *CG3209*, a de novo synthase of lysophosphatidic acid, is a genetic alteration that triggers a robust response of the animals to body size challenges in *Drosophila*. Loss of *CG3209* leads to a severe delay of development (extended development time), slow growth and resultant small-sized animals during the larval stages, but results in normal-sized adult flies. The robust body size adjustment of the *CG3209* mutant is likely achieved by corresponding changes in ecdysone and insulin signaling, which is also manifested by compromised food intake [163]. The reduced ecdysone signal in *CG3209* mutants may be responsible for the remarkable extension of developmental time; it is possible that the nutrient demand is altered for the *CG3209*

mutants. In Yan et al. (2015)[163] work, food intake was greater during the early larval stages and was reduced during the wandering stage for control animals. In contrast, the food intake for the *CG3209* mutants was less than the control during the second instar larval stage, and was similar for the early third instar larvae, whereas more food intake was detected for the mutants than the control at the wandering stage. Loss of *CG3209* reduces the rate of food intake during early larval stages. This reduced food intake is consistent with the observations of decreased insulin signaling and prolonged developmental time, which is in principle essential for ensuring normal-sized adults by the robust readjustment of developmental programs [163].

III.2.4. yki (yorkie)

The gene *yki* also known as *yorkie* is located in Muller's element C and shows signs of positive selection in all 3 of its known *D. melanogaster* transcripts according to bpositive database BP2016000001 [110]. The known *D. virilis* homolog is named *GJ20440*. *yki* is present in GO terms related to development such as "regulation of growth"; "cell proliferation"; "negative regulation of apoptotic process"; "anatomical structure"; "developmental process"; "positive regulation of growth" among others. It also appears in GO term: "*hippo* signaling". This pathway is known to regulate cell proliferation and maintain tissue homeostasis during development [180].

III.2.5. hpo (hippo)

The gene *hpo* also known as *hippo* is located on Muller's element C, and doesn't show signs of positive selection in all 2 of its transcripts according to bpositive database BP2016000001 [110]. The known *D. virilis* homolog is named *GJ21419*. *hpo* is present in GO terms related to development such as "eye development"; "apoptotic process"; "negative regulation cell proliferation"; "positive regulation of apoptotic signaling pathway"; "organ growth" among others. It is the main gene of the *hippo* signaling pathway that as previously mentioned is known to regulate cell proliferation and maintain tissue homeostasis during development. The interesting complex interaction between *hpo* and *yki* and its role in the *hippo* signaling pathway is mentioned below.

During normal development, the number of cells in growing tissues is controlled through regulation of cell proliferation and apoptosis [195-197]. The *Hippo* pathway controls organ size by multiple mechanisms [198-203], coordinately regulating these processes *hpo* signaling restricts tissue growth by promoting termination of cell proliferation and by stimulating apoptosis

during development, mechanisms that ultimately regulate the transcriptional co-activator *yki* (Fig. 8). At the centre of the *Hippo* signaling pathway is a core kinase cascade composed of two Serine/Threonine-specific kinases *hpo* [198-199, 204] and *Wts* (Warts) [205-206] and their adaptor proteins *Salvador* [204, 207] and *Mob* as tumour suppressor (*Mats*) [208]. The kinase cascade inactivates the oncoprotein *yki* [202] which acts as a transcriptional co-activator of the *Hippo* pathway, and responds to multiple upstream regulators like the FERM-domain containing proteins *Ex* (Expanded) and *Mer* (Merlin) [209] the WW- and C2-domain containing protein *Kibra* [210-211] and the single-pass trans-membrane proteins *Crb* (Crumbs) [212-213] and *Fat* [214]. *yki* partners with at least three transcription factors, *Scalloped*, *Homothorax* and *Teashirt*, to regulate transcription of genes regulating cellular growth (*myc*, *bantam*), cell cycle progression (*cyclinB*, *cyclinD*, *cyclinE*, *e2F1*) and inhibition of apoptosis (*diap1*). With its multiple inputs and targets, *Hippo* signaling has emerged as a pleiotropic pathway that acts in a context-dependent and tissue-specific manner to regulate organ size. *Wts* together with its cofactor *Mats* suppresses the transcriptional activity of *yki*, possibly through phosphorylation [202, 208]. Alternatively, *hpo*, *Wts* and *Ex* can directly interact with *yki* in a phosphorylation-independent manner leading to cytoplasmic sequestration and inhibition of *yki* activity. Hyperactivation of *Hippo* signaling in flies, by overexpression of *hpo*, *Sav*, *Mats*, and *Wts* has been shown to induce cell death and form smaller organs [198, 204].

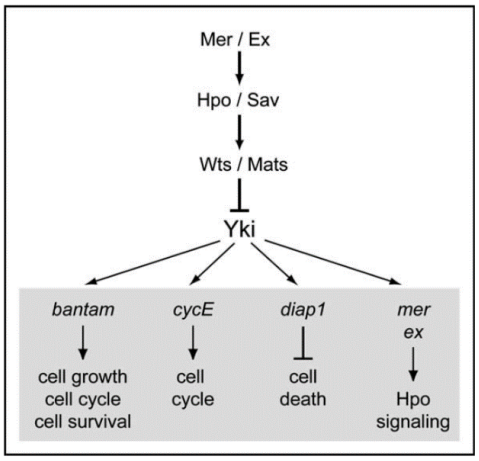


Figure 8- Model for *hippo* signaling pathway. The proteins (*Ex*) and (*Mer*) act upstream of the (*hpo*) kinase, which in turn activates the (*Wts*) kinase with help from the (*Sav*) adaptor protein. *Warts*, together with its cofactor, *Mats*, suppresses the activity of the transcriptional coactivator *yki*. *yki* drives the expression of the downstream target genes *bantam*, *cyclin E*, *diap1*, *ex*, and *mer*. The expression of these effector genes is correlated with cell growth, cell-cycle progression, and cell death. (adapted from [215])

On the other end, *Ex*, *Mer*, *Hpo*, *Sav*, *Wts*, and *Mats* mutants tissues develop into severely overgrown adult structures because mutant cells continue to proliferate after organs reach their normal size and because they are resistant to proapoptotic signals that normally eliminate extra cells. Overexpression of *yki* causes severe overgrowths that resemble the loss-of-function phenotypes of the other pathway members, whereas cells mutant for *yki* grow poorly [202]. *yki* is required for the overgrowth of *hpo* or *wts* mutant cells in vivo, and it was thus postulated that *yki* mediates most if not all of the growth effects of *hpo* signaling, presumably by driving the expression of transcriptional target genes [202]. When *hpo* signaling is reduced, for example by mutations in *hpo*, *yki* is hypophosphorylated and active, driving the expression of target genes that promote cell proliferation and suppress apoptosis, resulting in tissue overgrowth. However, as mentioned above, in normal scenarios, the *hpo* and *Wts* kinases limit the growth of tissues through the inactivation of *yki* [202, 215].

III.2.6. Act79B (Actin 79B)

The gene *Act79B* also known as *Actin 79B* is located in Muller's element D, and doesn't show signs of positive selection in both 2 of its known *D. melanogaster* transcripts according to bpositive database BP2016000001 [110]. The known *D. virilis* homolog is named *ActD1*. *Act79B* is present in GO terms such as "structural constituent of cytoskeleton"; "cytoskeleton organization" among others. It is also considered as a potential molecular marker of ontogenetic delay in *D. americana* by Reis et al. (2015) [181].

Actin is a ubiquitous and highly conserved eukaryotic protein required for cell motility and locomotion, whom in many organisms is encoded by multigene families in which individual isoforms probably perform cell-specific functions [216]. The four muscle actin genes of the insect *D. virilis* (closely related to *D. americana* and *D. novamexicana*) have strong similarities with their homologs in *D. melanogaster*; intron locations are conserved and there are few amino acid differences between homologs. The *D. virilis* gene homologous to *Act79B* is *ActD1* and it's specific to adult stages [217]. The predominant expression of *ActD1* is found in the TDT (tergal depressor of the trochanter) muscle. *ActD1* is also expressed in some other tubular muscles of the legs [218]. *ActD1* expression level is highly correlated with lifespan extension, given that LS is extended both in diapausing and in non-diapausing females reared under diapause inducing conditions. Reis et al. (2015) [181] hypothesized that flies under those conditions should have a gene expression pattern that is compatible with flies younger than their chronological age. Both in *D. melanogaster* and *D. virilis*, *Actin79B* (*ActD1* in *D. virilis*) expression levels drop dramatically with adult age. In *D. americana*, *ActD1* expression is highest in newborn females followed by diapausing individuals, non-diapausing individuals reared under diapause inducing

conditions and finally in individuals reared under 12L:12D at 25°C. Thus, also as expected, the *ActD1* expression levels accurately reflect the observed differences in LS. Suggesting that the expression of this gene can be used as a molecular marker of biological age [181]. Which will provide further insights into wether a fly who takes more time to develop has a different biological age than a fly who takes less time to develop, when comparing flies with the same chronological age.

The expression of these six candidate genes was studied using RT-PCR according to the method presented in Material and Methods and the results are shown and discussed below.

III.3. Candidate Genes Expression

Expression of *Sin3A*, *sm*, *CG3209*, *yki*, *hpo,* and *Act79B* was studied using RT-PCR. A total of 12 different conditions were set for three different strains. *D. americana* W11 flies with 15 days of development (from egg to adult) at 4 days and 8 days of life, and W11 flies with 16 days of development at 4 days and 8 days of life. *D. americana* SF12 flies with 15 days of development at 4 days and 8 days of life, and SF12 flies with 16 days of development at 4 days and 8 days of life. *D. novamexicana* NOVA00 flies with 16 days of development at 4 days and 8 days of life and NOVA00 flies with 17 days of development at 4 days and 8 days of life, as described in Material and Methods.

The full detailed results of all RT-PCR CT (cycle threshold) values for all 12 different conditions in the three different strains and the six candidate genes are displayed in Supplementary Tables 3-8. The data is shown in a similar way for each one of the six candidate genes, identifying the strain and condition studied and the biological replicate (from A to C). CT values for the candidate gene: the two different technical replicates CT values (CT.1 and CT.2) and the mean of these are also presented. The reference gene used (*RPL32* for all candidate genes) and its corresponding technical replicates CT values (CT.1 and CT.2) and the mean of these. The last column shows ΔCT (variation of cycle threshold) values calculated for the difference between mean CT values of the candidate gene and the reference gene. ΔCT values were calculated for each biological replicate.

For each gene, four different comparisons were made for each studied strain, presented using the following nomenclature. The first number corresponds to the days of development and the second number correspons to the days of life separated by "x". For instance, "15x4" represents flies with 15 days of development at 4 days of life.

The mean of the 3 biological replicates CT for each studied condition for each gene in a specific strain is presented in detail in Figures 9-14. For *Sin3A*, in Figures 9a, 9b and 9c, for *sm* in Figures 10a, 10b and 10c, for *CG3209* in Figures 11a, 11b and 11c, for *yki* in Figures 12a, 12b and 12c, for *hpo* in Figures 13a, 13b and 13, and for *Act79B* in Figures 14a, 14b and 14c. Furthermore, two-sample two-sided Student's t-test was applied to the data to address if there are statistically significant differences in gene expression, the results are displayed with significance bars (Figures 9-14) and the t-test values are presented for all 12 different comparisons for the 6 candidate genes (Table 13).

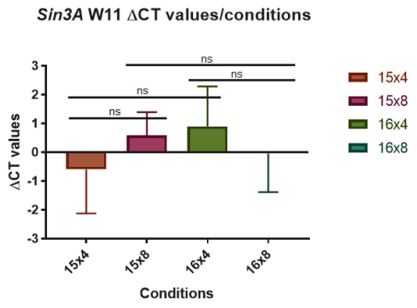


Figure 9a - *Sin3A* W11 ΔCT values per condition; Bars display t-test results for the comparisons made

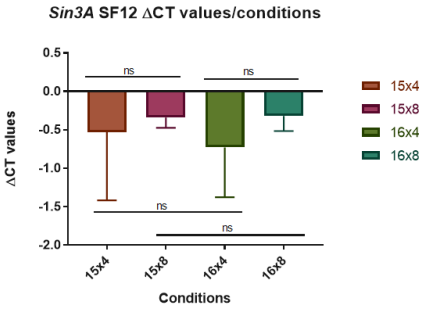


Figure 9b - *Sin3A* SF12 ΔCT values per condition; Bars display t-test results for the comparisons made

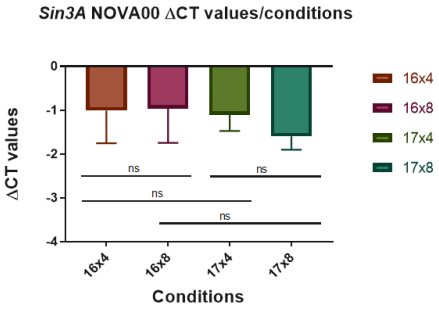


Figure 9c - *Sin3A* NOVA00 ΔCT values per condition; Bars display t-test results for the comparisons made

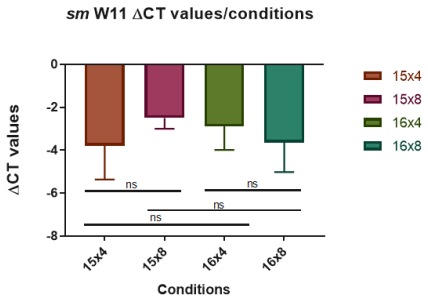


Figure 10a - *sm* W11 ΔCT values per condition; Bars display t-test results for the comparisons made

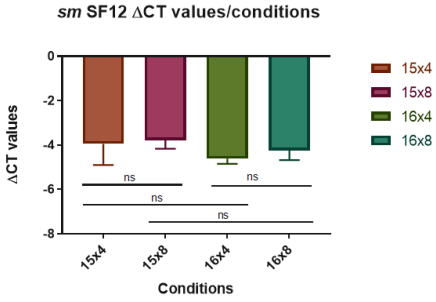


Figure 10b – *sm* SF12 ΔCT values per condition; Bars display t-test results for the comparisons made

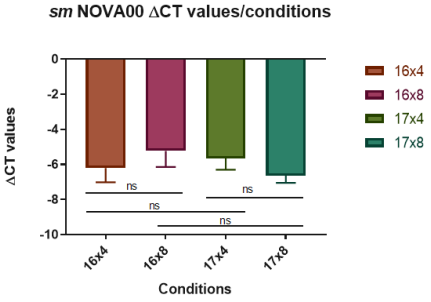


Figure 10c - *sm* NOVA00 ΔCT values per condition; Bars display t-test results for the comparisons made

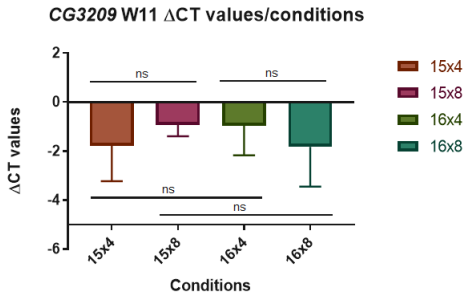


Figure 11a – *CG3209* W11 ΔCT values per condition; Bars display t-test results for the comparisons made

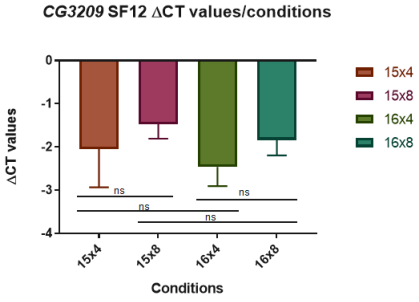


Figure 11b - *CG3209* SF12 ΔCT values per condition Bars display t-test results for the comparisons made

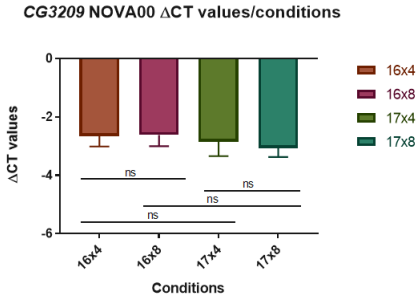


Figure 11c – *CG3209* NOVA00 ΔCT values per condition; Bars display t-test results for the comparisons made

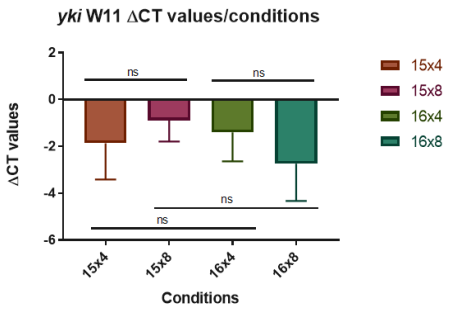


Figure 12a - *yki* W11 ΔCT values per condition; Bars display t-test results for the comparisons made

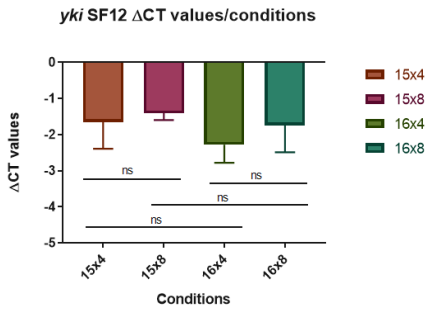


Figure 12b – *yki* SF12 ΔCT values per condition; Bars display t-test results for the comparisons made

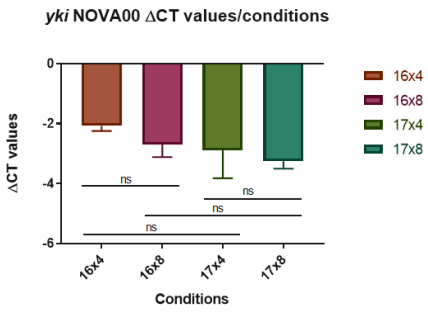


Figure 12c - *yki* NOVA00 ΔCT values per condition; Bars display t-test results for the comparisons made

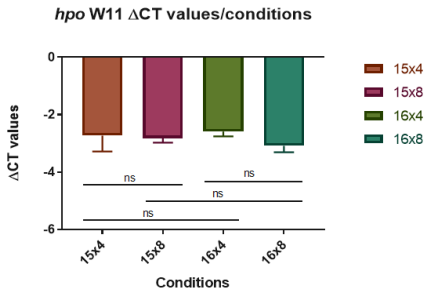


Figure 13a - *hpo* W11 ΔCT values per condition; Bars display t-test results for the comparisons made

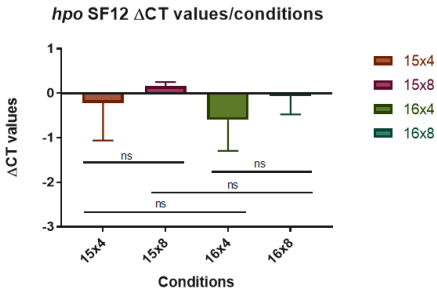


Figure 13b - *hpo* SF12 ΔCT values per condition; Bars display t-test results for the comparisons made

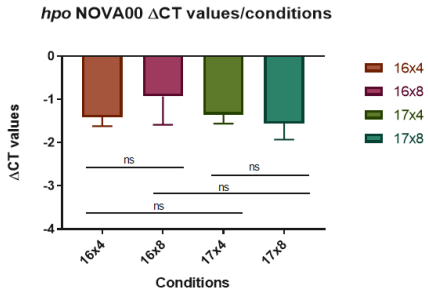


Figure 13c – *hpo* NOVA00 ΔCT values per condition; Bars display t-test results for the comparisons made

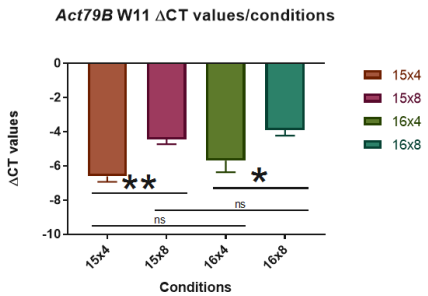


Figure 14a - *Act79B* W11 ΔCT values per condition; Bars display t-test results for the comparisons made

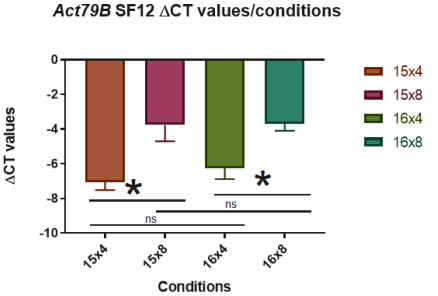


Figure 14b – *Act79B* SF12 ΔCT values per condition; Bars display t-test results for the comparisons made

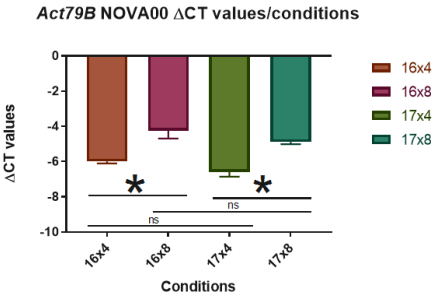


Figure 14c - *Act79B* NOVA00 ΔCT values per condition; Bars display t-test results for the comparisons made

Table 13 - Results of t-test per strain, per comparison and per gene studied. Highlights show significant results.

	Comparison	*Sin3A*	*sm*	*CG3209*	*hpo*	*yki*	*Act79B*
W11	**15x4 v 16x4**	0.3114	0.4970	0.5290	0.6458	0.7187	0.2487
	15x8 v 16x8	0.3343	0.3566	0.3283	0.2464	0.0816	0.2302
	15x4 v 15x8	0.2181	0.3609	0.2926	0.7550	0.5474	**0.0046**
	16x4 v 16x8	0.4202	0.5064	0.5219	0.1557	0.3811	**0.0150**
SF12	**15x4 v 16x4**	0.8438	0.4332	0.6220	0.7079	0.4755	0.3080
	15x8 v 16x8	0.9212	0.3323	0.3815	0.3777	0.4232	0.8841
	15x4 v 15x8	0.7112	0.8489	0.2196	0.5506	0.6455	**0.0449**
	16x4 v 16x8	0.3546	0.2692	0.2340	0.3190	0.2358	**0.0417**
NOVA00	**16x4 v 17x4**	0.7283	0.5362	0.1363	0.5194	0.2252	0.1059
	16x8 v 17x8	0.2531	0.0748	0.3409	0.1173	0.1616	0.1879
	16x4 v 16x8	0.9491	0.1129	0.9411	0.4330	0.1426	**0.0221**
	17x4 v 17x8	0.3339	0.2332	0.2645	0.5659	0.6394	**0.0133**

For *Sin3A*, *sm*, *CG3209*, *hpo*, and *yki*, there are no significant differences between the conditions tested for all 12 comparisons in the 3 different strains (Fig. 9-13 and Table 13). These results suggest that there are no differences in the expression of these genes between 15 and 16 (or 16 and 17) days of development and that there are no differences in the expression of the genes at 4 and 8 days of life.

For *Act79B*, there are no significant differences, between the conditions tested, for comparisons of the same day of life in different development times in all the 3 different strains (Fig. 14 and Table 13). These results suggest that there are no differences in the expression of *Act79B* between 15 and 16 (or 16 and 17) days of development. However, the t-test value shows a significant difference in the expression of this gene at 4 and 8 days of life when comparing flies with the same development time. Suggesting that there is a decrease of the expression of *Act79B* between 4 and 8 days of life.

IV. Discussion

IV.1. Bioinformatic analysis of Candidate Genes

IV.1.1. Muller's Element C

The analyzes of 71 different GO (Gene Ontology) terms lists related with development suggests that no list containing a significant amount of genes (more than 200 genes) is enhanced with more genes than what was expected for the *D. melanogaster* chromosome 2R (Muller's Element C). The observed distribution of the genes in each chromosome follows closely the expected distribution. However, 6 GO term lists ("brain development", "muscle cell development"; "response to ecdysone"; "response to insulin"; "ectoderm development" and "muscle fiber development"), with less than 200 genes, showed significant differences in the observed amount of genes compared to the expected distribution. Furthermore, when we don't consider chromosome 4, the distribution shows only 5 GO term lists with differences in observed results, suggesting that "brain development" is enhanced by genes on chromosome 4. Moreover, "muscle fiber development", a small list of 12 genes shows enhancement of genes in chromosome 2R, a result we suggest as not biologically significant, and a mere result of a small list of genes being used. The analyzes of these lists suggests that for most general biological processes there isn't an enhancement of genes in a specific chromosome, and that for specific biological processes such as "response to ecdysone" and "response to insulin" there seems to be an enhancement of genes in certain chromosomes, for "response to ecdysone" on chromosome 3L (Muller's element D) and X (Muller's element A) and for "response to insulin" on chromosome 3L (Muller's element D) and 3R (Muller's element E).

Therefore, looking at the results of Reis et al. (2014) [1], who found nine molecular markers showing significant associations with development time being eight of those on Muller's element C, such association cannot be due to a biased distribution of development genes on the genome. We thus suggest that the reported association is due to a few developmental genes of large effect on Muller's element C.

IV.1.2. Signs of positive selection

For such an important and vital process for organisms as development, it is theoretically unlikely to find a high percentage of genes showing signs of positive selection as genes involved in early developmental processes are expected to be under strong negative selection to prevent deleterious cascading effects [140]. Moreover, von Baer's third law, suggests that developmental conservation among related species implies that early stages of development are the most refractory to change and studies in *Drosophila* species have indicated that mid-embryogenesis is the stage most refractory to evolutionary change [140]. However, positively selected genes are found in all embryonic stages, even during the period with the highest developmental constraint, which is evidenced by our analyzes that shows many genes with signs of positive selection in GO terms associated with development processes. Therefore, even in functional groups of genes evolving under strong negative selection, there are many genes showing evidence of positive selection at the amino acid level.

When identifying genes showing signs of positive selection according to the bpositive database (http://bpositive.i3s.up.pt/) [110], the validation of these genes as genes showing signs of positive selection is tricky, and it is possible that some of the genes are false positives, as they can result in not optimal alignments or few sequences used.

Furthermore, in our analyses of positively selected genes. For the four candidate genes chosen that followed all our criteria, two of the three scenarios where a gene has transcripts who show signs of positive selection and other transcripts who don't show signs of positive selection, the ones showing these signs are transcripts with the lengthiest coding sequences. Suggesting that these transcripts have an additional region of the sequence that is being used in a specific tissue and/or specific lifestage.

IV.2. Candidate Genes Expression

We choose 4 candidate genes showing signs of positive selection and located on Muller's element C, for gene expression analyses. *Sin3A* and *sm*, who are also present under GO term: "determination of adult lifespan", which shows correlations with development time (DT) [1]. *CG3209*, who is also present in the GO term: "ecdysone receptor-mediated signaling pathway" and GO term: "response to insulin", and there is a reported suggestion that genes belonging to the insulin and ecdysone signaling pathways explain development time variations in *D. americana* [1]. And *yki*, an important gene belonging to the *Hippo* signaling pathway. This pathway is known

to regulate cell proliferation and maintain tissue homeostasis during development [180]. We selected two other candidate genes, *hpo*, who is the center of Hippo signaling pathway. And lastly, *Act79B*, whose expression levels can be used as a molecular marker of ontogenetic delay and are correlated with lifespan in *D. americana* [110].

Act79B shows significant differences, according to t-test (Figures 14a, 14b and 14c and Table 13), for the ΔCT (variation of cycle threshold) values when comparing 4 and 8 days of life for the same development time in the 3 studied strains, suggesting that the expression of this gene drops in adult life. This result is in accordance with several data available for *Act79B* expression. As reported previously [218], *Act79B* expression levels drop with age in both *D. melanogaster* and *D. virilis* adults.

Moreover, *Act79B* shows no differences in expression in flies with different development times at the same day of life in the 3 studied strains (Figures 14a, 14b and 14c and Table 13). Considering that *ActinD1* (the *virilis* homolog of *Act79B*) expression levels can be used as a marker of ontogenetic delay [181], this results suggests that flies with different development time are not showing differences in ontogenetic development. Specifically, a fly with 16 development days is not showing significant ontogenetic delay when compared to a fly with 15 development days. Therefore, we argue that differences in development time are not affecting the ontogenetic development of the flies studied.

None of the other candidate genes (*Sin3A*, *sm*, *CG3209*, *yki* and, *hpo*) showed significant differences, according to t-test (Figures 9-13 and Table 13), for the ΔCT values for the comparisons made on the same days of adult life for different development times for the 3 studied strains. This result is unexpected at first if we consider the already reported involvement of ecdysone and insulin signaling pathways and the association/correlation between development time and lifespan, we expected to find differences in the expression of the candidate genes between different development times.

The main criteria for the search of candidate genes were set for previous suggestions for association with developmental processes, chromosomal presence (Muller's element C) and evidence for positive selection in the amino acid sequence. Reis et al. (2014)[1], performed an F2 association study for Development Time (DT) and Lifespan (LS). For the *D. americana* H5xW11 F2 association study they verified nine molecular markers showing significant associations with development time, eight of which were present on Muller's element C. This study suggests a strong connection between Muller's element C genes and development time in *D. americana*. As already mentioned, 5 of the 6 studied candidate genes are located in Muller's element C, thus it is surprising that none shows different levels of expression for the selected development times. It could be argued that other candidate genes in Muller's element C have a bigger impact in development time and that it shows little recombination.

Four of the candidate genes (*Sin3A*, *sm*, *CG3209*, and *yki*) are considered "positively selected", however it is important to note that *Sin3A* shows signs of positive selection in 5 of its 10 know transcripts, *sm* in 3 of its 23 known transcripts, *CG3209* in 2 of its 4 known transcripts and *yki* in all of its 3 transcripts. Only *yki* has all transcripts classified as "positively selected". Taking this into account, it is important to remember that the relative abundance of each gene transcript in the extracted RNA could result in testing the expression of the more abundant transcripts. These could be transcripts who don't show signs of positive selection and show no differences in expression levels. Indeed, the primers were designed in order to amplify all transcripts (for regions of the sequence who are similar in all transcripts). It is also possible that the different transcripts of a gene are isoforms who are being transcribed in specific tissues. Nevertheless, the RNA extraction was made in whole fly bodies and this problem is reduced. The transcripts can also be stage-specific transcripts, present only on larval or adult stages, and considering that we snap froze flies at 4 and 8 days of life, we are extracting the RNA that is being transcribed at these specific stages of the adult life.

It can be argued that the genes that are associated with development time and play important roles during development, could have a more impactful expression in larval stages than in adult life stages, having an impact on the developmental time variation during this stage. It would, therefore, be interesting to study the expression of these genes during larval stages. However, that would be impossible, since if we use larvae for studying development time, we wouldn't be capable of knowing the development time phenotypes of the studied larvae. Moreover, even if we use an isogenic strain with characterized development time, intraspecific variation should be taken into account. It is highly difficult to guarantee that two different larvae are in the same development stage, and furthermore, it is complicated to synchronize different larvae to a specific developmental stage. Therefore, we studied development time variations in adult stage taking into account the assumption that gene expression is correlated between larvae and adult flies.

By snap freezing the flies at 4 and 8 days of life we verified whether or not there is an increase or decrease of expression of these genes during adult life, especially because after 4 days of life the expression of the genes are usually stabilized. The expression of the candidate genes could have been different if other different days of adult life were chosen to snap freeze samples, but it is unknown what are the best days to assay differences in gene expression. Using newly born flies should be done with caution, as the gene expression in the early stages of adult life is usually remarkably variable. Therefore, biological samples of the same condition could have produced significantly different values of CT, due to few hours of life differences that would impact gene expression.

Moreover, flies were obtained for 15 and 16 days of development in the *D. americana* strains and for 16 and 17 days of development in the *D. novamexicana* strain, so the comparisons were performed between the specific development times. Choosing other development times, 14,

15, 16, 17 or 18 for *D. americana* and *D. novamexicana* would augment the amount of possible comparisons to be done. It would, however, be arduous and highly time-demanding to obtain the necessary amount of male flies to eventually snap freeze flies with different days of adult life. Considering the difficulties associated with RNA extraction, RNA integrity assays and the necessary quality of samples for RT-PCR, obtaining the necessary amount of flies for the less frequent development days could be incredibly challenging. Regardless of this, the difference an extra day makes in terms of development in an organism with such a short life cycle like *Drosophila* is still considerably relevant and the data obtained is still remarkable.

Lastly, for the chosen genes, flies with different development days do not show expression differences. Flies with different development time, don't show differences in the expression of *Act79B* either, a result that suggests that variations in development time are not correlated with ontogenetic development. Therefore, taking this experiment into account we can argue that it is not unexpected that the candidate genes are not showing differences in expression in flies with different developmental times. *Act79B* expression levels show that there isn't an ontogenetic delay in flies with increased development time. Flies with different development times seem to be in the same developmental stage at 4 and 8 days of life. Therefore, it can be argued the expression of genes is the same in this stage, as this flies are in the same ontogenetic development stage.

V. Conclusion

Our results suggest that no GO (Gene Ontology) term list related with development is enhanced with more genes than expected for the chromosome 2R. Although Reis et al. (2014) [1], suggest a connection between the Muller's element C genes and development time, we hypothesize that a small number of genes in Muller's element C should account for the majority of the correlation.

Our results for the expression of *Act79B* in flies with the same development time and different days of life further suggest that the expression of this gene is decreasing in adult life. The study of the expression of *Act79B* in *D. americana* W11 and SF12 and *D. novamexicana* NOVA00 shows that there are no significant differences between the expression of this gene in flies with different development times on the same day of life. Considering that *Act79B* expression can be used as a marker of ontogenetic delay in *Drosophila,* the observation of our results suggests that flies with 15 and 16 (or 16 and 17) days of development are in the same ontogenetic development stage at the days of adult life used for testing expression.

For the other remaining 5 candidate genes (*Sin3A*, *sm*, *CG3209*, *yki*, and *hpo*) we found no significant differences in expression levels when comparing flies from the studied strains *D. americana* W11 and SF12 and *D. novamexicana* NOVA00 at different days of life and when comparing flies with different development times. The chosen genes are involved in the *ecdysone*, *insulin,* and *hippo* signaling pathways, all of which play an important role in developmental processes. Thus it is surprising that these genes show no differences in expression when comparing flies with different development times.

However, when we consider the results for *Act79B* (for flies with different development times there are no significant differences in expression) and taking into account that its expression can be used as a marker of ontogenetic delay in *Drosophila*, the finding that none of the five other candidate genes show no difference in expression in flies at the same day of life and different development times might be explained by the fact that there seem to be no differences in ontogenetic development in flies with 15 and 16 (or 16 and 17) days of development.

Further studies about these genes should be done with different experimental setups, abording different development times, different days of life to test for gene expression and different techniques. Moreover, although highly difficult and theoretically impossible, the efforts into studying development time in larvae should be exploited.

Much remains unclear about what causes a fly to have 15 days of development and what causes another to have 16 days of development. As well as the causes, also the consequences aren't clear, although reported association with lifespan has been shown before, our findings show that there isn't an association between development time variations and ontogenetic delay.

VI. References

[1] M. Reis, I. Páscoa, H. Rocha, B. Aguiar, C. P. Vieira, and J. Vieira, "Genes belonging to the *insulin* and *ecdysone* signaling pathways can contribute to developmental time, lifespan and abdominal size variation in *Drosophila americana*" *PLoS One*, vol. 9, no. 1, 2014.

[2] J. A. Mossman, L. M. Biancani, C.-T. Zhu, and D. M. Rand, "Mitonuclear Epistasis for Development Time and Its Modification by Diet in *Drosophila*" *Genetics*, vol. 203, no. 1, pp. 463–484, May 2016.

[3] B. J. Zwaan, R. Bijlsma, and R. F. Hoekstra, "On the developmental theory of ageing. I. starvation resistance and longevity in *Drosophila melanogaster* in relation to pre-adult breeding conditions." *Heredity (Edinb).*, vol. 66 (Pt 1), pp. 29–39, Feb. 1991.

[4] M. F. Wangler, Y. Hu, and J. M. Shulman, "*Drosophila* and genome-wide association studies: a review and resource for the functional dissection of human complex traits" *Dis. Model. Mech.*, vol. 10, no. 2, pp. 77–88, 2017.

[5] G. Dos Santos *et al.*, "*FlyBase*: Introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations" *Nucleic Acids Res.*, vol. 43, no. D1, pp. D690–D697, 2015.

[6] G. H. Millburn, M. A. Crosby, L. S. Gramates, and S. Tweedie, "FlyBase portals to human disease research using *Drosophila* models" *Dis. Model. Mech.*, vol. 9, no. 3, pp. 245–252, 2016.

[7] B. P. Berman *et al.*, "Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome" *Proc. Natl. Acad. Sci.*, vol. 99, no. 2, pp. 757–762, 2002.

[8] D. A. Papatsenko *et al.*, "Extraction of Functional Binding Sites from Unique Regulatory Regions : The *Drosophila* Early Developmental Enhancers Extraction of Functional Binding Sites from Unique Regulatory Regions" *Genome Res.*, pp. 470–481, 2002.

[9] M. F. Wangler, S. Yamamoto, and H. J. Bellen, "Fruit flies in biomedical research" *Genetics*, vol. 199, no. 3, pp. 639–653, 2015.

[10] S. P. Voght, M. L. Fluegel, L. A. Andrews, and L. J. Pallanck, "*Drosophila NPC1b* Promotes an Early Step in Sterol Absorption from the Midgut Epithelium" *Cell Metab.*, vol. 5, no. 3, pp. 195–205, Mar. 2007.

[11] K. Venkatachalam, A. A. Long, R. Elsaesser, D. Nikolaeva, and C. Montell, "Motor deficit in a *Drosophila* model of mucolipidosis type IV due to defective clearance of apoptotic cells" *Cell.*, vol. 135, no. 5, pp. 838–851, 2008.

[12] H. Weavers *et al.*, "The insect nephrocyte is a podocyte-like cell with a filtration slit diaphragm" *Nature*, vol. 457, no. 7227, pp. 322–326, 2009.

[13] S. B. Diop, J. Bisharat-Kernizan, R. T. Birse, S. Oldham, K. Ocorr, and R. Bodmer, "*PGC-1/spargel* counteracts high fat diet-induced obesity and cardiac lipotoxicity downstream of *TOR* and *Brummer ATGL* lipase" *Cell Rep.*, vol. 10, pp. 1572–1584, 2015.

[14] M. Jaiswal, H. Sandoval, K. Zhang, V. Bayat, and H. J. Bellen, "Probing Mechanisms That Underlie Human Neurodegenerative Diseases in *Drosophila*" *Annu. Rev. Genet.*, vol. 46, no. 1, pp. 371–396, 2012.

[15] D. Lessing and N. M. Boninin, "Maintaining the brain: insight into human neurodegeneration from *Drosophila melanogaster* mutants" *Nat. Rev. Genet.*, vol. 10, pp. 359–370, 2009.

[16] J. M. Shulman, L. M. Shulman, W. J. Weiner, and M. B. Feany, "From fruit fly to bedside: translating lessons from *Drosophila* models of neurodegenerative disease." *Curr. Opin. Neurol.*, vol. 16, no. 4, pp. 443–449, 2003.

[17] E. Romero *et al.*, "Suppression of Neurodegeneration and Increased Neurotransmission Caused by Expanded Full-Length Huntingtin Accumulating in the Cytoplasm" *Neuron*, vol. 57, no. 1, pp. 27–40, Jan. 2008.

[18] H. Tsuda *et al.*, "The AXH domain of *ataxin-1* mediates neurodegeneration through its interaction with *Gfi-1/senseless* proteins" *Cell*, vol. 122, no. 4, pp. 633–644, Aug. 2005.

[19] H. Tsuda *et al.*, "The Amyotrophic Lateral Sclerosis 8 Protein *VAPB* Is Cleaved, Secreted, and Acts as a Ligand for *Eph* Receptors" *Cell*, vol. 133, no. 6, pp. 963–977, 2008.

[20] M. Vos *et al.*, "*Vitamin K2* Is a Mitochondrial Electron Carrier That Rescues *Pink1* Deficiency" *Science (80-.).*, vol. 6086, no. January, pp. 1306–1310, 2012.

[21] G. G. Neely *et al.*, "A genome-wide *Drosophila* screen for heat nociception identifies *α2δ3* as an evolutionary-conserved pain gene" *Cell.*, vol. 143, no. 4, pp. 628–638, 2010.

[22] G. G. Neely *et al.*, "A global in vivo *Drosophila* RNAi screen identifies *NOT3* as a conserved regulator of heart function" *Cell.*, vol. 141, no. 1, pp. 142–153, 2010.

[23] J. A. Pospisilik *et al.*, "*Drosophila* Genome-wide Obesity Screen Reveals *Hedgehog* as a Determinant of Brown versus White Adipose Cell Fate" *Cell*, vol. 140, no. 1, pp. 148–160, 2010.

[24] S. Yamamoto *et al.*, "A *Drosophila* Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases" *Cell*, vol. 159, no. 1, pp. 200–214, Sep. 2014.

[25] Drosophila 12 Gemomes Consortium, "Evolution of genes and genomes on the *Drosophila* phylogeny" *Nature*, vol. 450, no. 7167, pp. 203–218, 2007.

[26] N. A. Fonseca *et al.*, "*Drosophila americana* as a model species for comparative studies

on the molecular basis of phenotypic variation" *Genome Biol. Evol.*, vol. 5, no. 4, pp. 661–679, 2013.

[27] R. Morales-Hojas and J. Vieira, "Phylogenetic Patterns of Geographical and Ecological Diversification in the Subgenus *Drosophila*" *PLoS One*, vol. 7, no. 11, 2012.

[28] G. Bosco, P. Campbell, J. T. Leiva-Neto, and T. A. Markow, "Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species" *Genetics*, vol. 177, no. 3, pp. 1277–1290, 2007.

[29] J. R. Manak *et al.*, "Biological function of unannotated transcription during the early development of *Drosophila melanogaster*" *Nat. Genet.*, vol. 38, pp. 1151–1158, 2006.

[30] A. Heger and C. P. Ponting, "Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes" *Genome Res.*, vol. 17, no. 12, pp. 1837–1849, 2007.

[31] R. H. Waterston *et al.*, "Initial sequencing and comparative analysis of the mouse genome" *Nature*, vol. 420, no. 6915, pp. 520–562, 2002.

[32] B. Negre, J. Ranz, F. Casals, M. Caceres, and A. Ruiz, "A New Split of the *Hox* Gene Complex in *Drosophila*: Relocation and Evolution of the Gene *labial*", Mol. Biol. Evol., vol. 20, no. 12, pp. 2042-2054, 2004.

[33] G. Von Allmen *et al.*, "Splits in fruitfly *Hox* gene complexes" *Nature*, vol. 380, no. 6570, p. 116, 1996.

[34] B. Negre and A. Ruiz, "*HOM-C* evolution in *Drosophila*: is there a need for *Hox* gene clustering?" *Trends in Genetics*, vol. 23, no. 2. Elsevier Current Trends, pp. 55–59, 01-Feb-2007.

[35] L. H. Throckmorton, "The phylogeny, ecology and geography of *Drosophila"*, *Handbook o. New York: Plenum Publishing Corporation*, 1975.

[36] P. M. O'Grady and T. A. Markow, "Phylogenetic taxonomy in *drosophila*: Problems and prospects" *Fly (Austin).*, vol. 3, no. 1, pp. 10–14, 2009.

[37] T. A. Markow and P. O'Grady, "Reproductive ecology of *Drosophila*" *Funct. Ecol.*, vol. 22, no. 5, pp. 747–759, 2008.

[38] T. A. Markow and P. O'Grady, "*Drosophila*: A guide to species identification and use." *Elsevier*. London: Academic Press, 2006.

[39] J. Remsen and P. O'Grady, "Phylogeny of *Drosophilinae* (*Diptera*: *Drosophilidae*), with comments on combined analysis and character support" *Mol. Phylogenet. Evol.*, vol. 24, no. 2, pp. 249–264, Aug. 2002.

[40] C. A. M. Russo, N. Takezaki, and M. Nei, "Molecular Phylogeny and Divergence Times of *Drosophilid* Species." *Mol Biol Evol.*, vol. 12, pp. 391–404, 1995.

[41] T. A. Markow and P. M. O'Grady, "*Drosophila* biology in the genomic age" *Genetics*, vol. 177, no. 3, pp. 1269–1276, 2007.

[42] J. R. Powell, "Progress and Prospects in Evolutionary Biology: The *Drosophila* Model"

Oxford University Press New York, Oxford, p. 562, 1997.

[43] M. A. F. Noor, D. A. Garfield, S. W. Schaeffer, and C. A. Machado, "Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions" *Genetics*, vol. 177, no. 3, pp. 1417–1428, 2007.

[44] G. S. Spicer, "The Genetic Basis of a Species-Specific Character in the *Drosophila virilis* Species group" *Strain*, vol. 337, pp. 331–337, 1991.

[45] P. J. Wittkopp *et al.*, "Local adaptation for body color in *Drosophila americana*" *Heredity (Edinb).*, vol. 106, no. 4, pp. 592–602, 2011.

[46] A. M. Kulikov, A. I. Melnikov, N. G. Gornostaev, O. E. Lazebny, and V. G. Mitrofanov, "Morphometric analysis of male genitalia in sibling species of *Drosophila virilis*" *Russ. J. Genet.*, vol. 40, no. 2, pp. 125–138, 2004.

[47] B. C. Caletka and B. F. McAllister, "A genealogical view of chromosomal evolution and species delimitation in the *Drosophila virilis* species subgroup" *Mol. Phylogenet. Evol.*, vol. 33, no. 3, pp. 664–670, 2004.

[48] B. F. McAllister, "Genetic analysis of sex-chromosome arrangement in *Drosophila americana*: a laboratory exercise for undergraduate or advanced placement students." *Dros. Inf. Serv.*, vol. 84, pp. 227–234, 2001.

[49] B. F. McAllister, "Analysis of a hybrid zone in *Drosophila americana*: selective maintenance of a chromosomal cline." *Genome*, vol. 45, pp. 13–21, 2002.

[50] H. D. Stalker, "The inheritance of a subspecific character in the *virilis* complex of *Drosophila*" *Am. Nat.*, vol. 76, pp. 426–431, 1942.

[51] W. P. Spencer, "Subspecies, hybdris and speciation in *Drosophila hydei* and *Drosophila virilis*" *Am. Nat.*, vol. 74, pp. 157–179, 1940.

[52] Y. H. Ahmed-Braimah and A. L. Sweigart, "A single gene causes an interspecific difference in pigmentation in *Drosophila*" *Genetics*, vol. 200, no. 1, pp. 331–342, 2015.

[53] D. L. Stern, "A role of *Ultrabithorax* in morphological differences between *Drosophila* species" *Nature*, vol. 396, no. December, pp. 463–467, 1998.

[54] E. Sucena and D. L. Stern, "Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby" *Proc. Natl. Acad. Sci. USA*, vol. 97, no. 9, pp. 4530–4534, 2000.

[55] A. P. McGregor *et al.*, "Morphological evolution through multiple cis-regulatory mutations at a single gene" *Nature*, vol. 448, no. 7153, pp. 587–590, 2007.

[56] B. V Adrianov, S. Y. Sorokina, T. V Gorelova, and V. G. Mitrofanov, "Mitochondrial DNA polymorphism in natural populations of the *Drosophila virilis* species group" *Russ. J. Genet.*, vol. 39, no. 6, pp. 762–768, 2003.

[57] H. A. Orr and J. A. Coyne, "The genetics of postzygotic isolation in the *Drosophila virilis* group" *Genetics*, vol. 121, no. 3, pp. 527–537, 1989.

[58] A. L. Sweigart, “The genetics of postmating, prezygotic reproductive isolation between *Drosophila virilis* and *D. americana*” *Genetics*, vol. 184, no. 2, pp. 401–410, 2010.

[59] A. L. Sweigart, “Simple Y-autosomal incompatibilities cause hybrid male sterility in reciprocal crosses between *Drosophila virilis* and *D. americana*” *Genetics*, vol. 184, no. 3, pp. 779–787, 2010.

[60] R. Morales-Hojas, M. Reis, C. P. Vieira, and J. Vieira, “Resolving the phylogenetic relationships and evolutionary history of the *Drosophila virilis* group using multilocus data” *Mol. Phylogenet. Evol.*, vol. 60, no. 2, pp. 249–258, Aug. 2011.

[61] N. Sagga and A. Civetta, “Male-Female Interactions and the Evolution of Postmating Prezygotic Reproductive Isolation among Species of the *Virilis* Subgroup.” *Int. J. Evol. Biol.*, vol. 2011, p. 485460, 2011.

[62] Y. H. Ahmed-Braimah and B. F. McAllister, “Rapid Evolution of Assortative Fertilization between Recently Allopatric Species of *Drosophila*” *Int. J. Evol. Biol.*, vol. 2012, pp. 1–9, 2012.

[63] J. T. Patterson, “*Drosophila* and speciation.” *Science (80).*, vol. 95, pp. 153–9, 1942.

[64] L. H. Throckmorton, “The genetics and biology of *drosophila*” in *Academic Press*, New York, 1982, pp. 227–296.

[65] G. S. Spicer, “Reevaluation of the phylogeny of the *Drosophila virilis* species group (*Diptera*: *Drosophilidae*).” *Ann. Entomol. Soc. Am.*, vol. 85, pp. 11–25, 1992.

[66] H. Hilton and J. Hey, “DNA sequence variation at the period locus reveals the history of species and speciation events in the *Drosophila virilis* group” *Genetics*, vol. 144, no. 3, pp. 1015–1025, 1996.

[67] P. M. O’Grady, R. H. Baker, C. M. Durando, W. J. Etges, and R. DeSalle, “Polytene chromosomes as indicators of phylogeny in several species groups of *Drosophila*.” *BMC Evol. Biol.*, vol. 1, p. 6, 2001.

[68] G. S. Spicer and C. D. Bell, “Molecular phylogeny of the *Drosophila virilis* species group (*Diptera*: *Drosophilidae*).” *Ann Entomol Am*, vol. 95, pp. 156–161, 2002.

[69] A. M. Kulikov, A. I. Melnikov, N. G. Gornostaev, O. E. Lazebny, and V. G. Mitrofanov, “Morphological analysis of male mating organ in the *Drosophila virilis* species group: a multivariate approach” *J. Zool. Syst. Evol. Res.*, vol. 42, no. 2, pp. 135–144, 2004.

[70] L. Orsini, S. Huttunen, and C. Schlötterer, “A multilocus microsatellite phylogeny of the *Drosophila virilis* group.” *Heredity (Edinb).*, vol. 93, no. 2, pp. 161–165, 2004.

[71] D. L. Nurminsky, E. N. Moriyama, E. R. Lozovskaya, and D. L. Hartl, “Molecular Phylogeny and Genome Evolution in the *Drosophila virilis* Species Group: Duplications of the *Alcohol Dehydrogenase* Gene” *Mol. Biol. Evol.*, vol. 13, pp. 132–149, 1996.

[72] M. S. Ostrega, “Restriction Endonuclease Analysis of the Relatedness of *D. montana* and *D. virilis* Line” *Dros. Inf. Serv.*, vol. 61, pp. 132–133, 1985.

[73] J. T. Patterson and W. S. Stone, "Evolution in the genus *drosophila"*, *New York: The MacMillan Company. Science Education* vol. 38, no. 4. 1952.

[74] W. P. Spencer, "*Drosophila virilis americana*, a new sub-species" *Genetics*, vol. 23, pp. 169–170. 1938.

[75] R. D. Hughes, "An Analysis of the Chromosomes of the Two Sub-Species *Drosophila Virilis Virilis* and *Drosophila Virilis Americana*." *Genetics*, vol. 24, no. 6, pp. 811–834, 1939.

[76] Y. X. Fu and W. H. Li, "Statistical tests of neutrality of mutations" *Genetics*, vol. 133, no. 3, pp. 693–709, 1993.

[77] B. F. McAllister and B. Charlesworth, "Reduced sequence variability on the neo-Y chromosome of *Drosophila americana americana*" *Genetics*, vol. 153, no. 1, pp. 221–233, 1999.

[78] R. Morales-Hojas, C. P. Vieira, and J. Vieira, "Inferring the evolutionary history of *Drosophila americana* and *Drosophila novamexicana* using a multilocus approach and the influence of chromosomal rearrangements in single gene analyses" *Mol. Ecol.*, vol. 17, no. 12, pp. 2910–2926, 2008.

[79] J. Vieira, B. F. McAllister, and B. Charlesworth, "Evidence for selection at the *fused* locus of *Drosophila virilis*" *Genetics*, vol. 158, no. 1, pp. 279–290, 2001.

[80] B. F. McAllister, "Chromosomal and allelic variation in *Drosophila americana*: selective maintenance of a chromosomal cline." *Genome*, vol. 45, no. 1, pp. 13–21, 2002.

[81] B. F. McAllister, S. L. Sheeley, P. A. Mena, A. L. Evans, and C. Schlötterer, "Clinal distribution of a chromosomal rearrangement: A precursor to chromosomal speciation?" *Evolution (N. Y).*, vol. 62, no. 8, pp. 1852–1865, 2008.

[82] M. Reis, C. P. Vieira, R. Morales-Hojas, and J. Vieira, "An old *bilbo*-like non- LTR retroelement insertion provides insight into the relationship of species of the *virilis* group" *Gene*, vol. 425, pp. 48–55, 2008.

[83] J. C. Avise, "Phylogeography: History and Formation of Species" *Harvard Univeristy Press*, London, p. 447p, 2000.

[84] H. Hilton and J. Hey, "A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts." *Genet. Res.*, no. 70, pp. 185–194.

[85] J. C. Stephens and M. Nei, "Phylogenetic analysis of polymorphic DNA sequences of the *Adh* locus in *Drosophila melanogaster* and its sibling species." *J. Mol. Evol.*, no. 22, pp. 289–300, 1985.

[86] C. P. Vieira, A. Almeida, J. Daniel Dias, and J. Vieira, "On the location of the gene(s) harbouring the advantageous variant that maintains the X/4 fusion of *Drosophila americana*" *Genet. Res.*, vol. 87, no. 3, p. 163, 2006.

[87] N. A. Ayoub and S. E. Riechert, “Molecular evidence for Pleistocene glacial cycles driving diversification of a North American desert spider, *Agelenopsis aperta*” *Mol. Ecol.*, vol. 13, no. 11, pp. 3453–3465, 2004.

[88] T. A. Castoe, C. L. Spencer, and C. L. Parkinson, “Phylogeographic structure and historical demography of the western diamondback rattlesnake (*Crotalus atrox*): A perspective on North American desert biogeography” *Mol. Phylogenet. Evol.*, vol. 42, no. 1, pp. 193–212, 2007.

[89] D. J. Morafka, G. A. Adest, and L. M. Reyes, “Differentiation of North American deserts: a phylogenetic evaluation of a vicariance model.” in *Biogeography of Mesoamerica*, S. P. Darwin and A. L. Welden, Eds. New Orleans, Louisiana: Tulane Studies in Zoology and Botany Special Publication, 1992, pp. 195–226.

[90] “FlyBase Website” [Online]. Available: flybase.org. Data obtained from FB2017_02, released April 18, 2017

[91] J. T. Patterson, “The *virilis* group of *Drosophila* in Texas.” *Am. Nat.*, no. 75, pp. 523–529, 1941.

[92] T. C. Hsu, “Chromosomal variation and evolution in the *virilis* group of *Drosophila*” *Univ. Texas Publ.*, no. 5204, pp. 35–72, 1952.

[93] P. J. Wittkopp, B. L. Williams, J. E. Selegue, and S. B. Carroll, “*Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes” *Proc. Natl. Acad. Sci.*, vol. 100, no. 4, pp. 1808–1813, 2003.

[94] J. Vieira, C. P. Vieira, D. L. Hartl, and E. R. Lozovskaya, “A framework physical map of *Drosophila virilis* based on P1 clones: applications in genome evolution.” *Chromosoma*, vol. 106, no. 2, pp. 99–107, 1997.

[95] A. M. Cooley, L. Shefner, W. N. McLaughlin, E. E. Stewart, and P. J. Wittkopp, “The ontogeny of color: Developmental origins of divergent pigmentation in *Drosophila americana* and *D. novamexicana*” *Evol. Dev.*, vol. 14, no. 4, pp. 317–325, 2012.

[96] H. Hollocher, J. L. Hatcher, and E. G. Dyreson, “Evolution of abdominal pigmentation differences across species in the *Drosophila dunni* subgroup.” *Evolution (N. Y).*, vol. 54, pp. 2046–2056, 2000.

[97] J. A. Brisson, D. C. De Toni, I. Duncan, and A. R. Templeton, “Abdominal pigmentation variation in *Drosophila polymorpha*: geographic in the trait, and underlying phylogeography.” *Evolution (N. Y).*, vol. 59, pp. 1046–1059, 2005.

[98] J. E. Pool and C. F. Aquadro, “The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*” *Mol Ecol.*, vol. 16, no. 14, pp. 2844–2851, 2007.

[99] M. Telonis-Scott, A. A. Hoffmann, and C. M. Sgrò, “The molecular genetics of clinal variation: A case study of ebony and thoracic trident pigmentation in *Drosophila melanogaster* from eastern Australia” *Mol. Ecol.*, vol. 20, no. 10, pp. 2100–2110, 2011.

[100] D. R. Matute and A. Harris, “The influence of abdominal pigmentation on desiccation and ultraviolet resistance in two species of *Drosophila*” *Evolution (N. Y).*, vol. 67, no. 8, pp. 2451–2460, 2013.

[101] J. R. True, “Insect melanism: The molecules matter”, *Trends in Ecology & Evolution*, vol. 18, no. 12. 2003.

[102] P. J. Wittkopp and P. Beldade, “Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy”, *Semin Cell Dev Biol.*, vol. 20, no. 1. 2009.

[103] A. Kopp, “Metamodels and phylogenetic replication: A systematic approach to the evolution of developmental pathways” *Evolution (N. Y).*, vol. 63, no. 11, pp. 2771–2789, 2009.

[104] C. H. Waddington, “Body-colour genes in *Drosophila*.” *Proc. Zool. Soc. Lond. Ser. A* vol. 111, pp. 173–180, 1942.

[105] R. Levis, T. Hazelrigg, and G. M. Rubin, “Separable cis-acting control elements for expression of the white gene of *Drosophila*.” *EMBO J*, vol. 4, pp. 3489–3499, 1985.

[106] I. Dombeck and J. Jaenike, “Ecological genetics of abdominal pigmentation in *Drosophila falleni*: a pleiotropic link to nematode parasitism.” *Evolution (N. Y).*, vol. 58, pp. 587–596.

[107] N. Gompel, B. Prud’homme, P. J. Wittkopp, V. A. Kassner, and S. B. Carroll, “Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*.” *Nature*, vol. 433, pp. 481–487, 2005.

[108] S. Jeong, A. Rokas, and S. B. Carroll, “Regulation of body pigmentation by the abdominal-*B Hox* protein and its gain and loss in *Drosophila* evolution.” *Cell*, vol. 125, pp. 1387–1399, 2006.

[109] M. Rebeiz, J. E. Pool, V. A. Kassner, C. F. Aquadro, and S. B. Carroll, “Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population.” *Science (80)*, vol. 326, pp. 1663–1667, 2009.

[110] N. Vázquez *et al.*, “On the automated collection and sharing (B+ database) of data on adaptive amino acid changes” *11th International Conference on Practical Applications of Computational Biology & Bioinformatics*. 2017.

[111] D. A. Drummonds, J. D. Bloom, C. Adami, C. O. Wilke, and F. H. Arnold, “Why highly expressed proteins evolve slowly.” *Proc. Natl Acad. Sci. USA*, vol. 102, pp. 14338–14343, 2005.

[112] D. A. Drummond, A. Raval, and C. O. Wilke, “A single determinant dominates the rate of yeast protein evolution.” *Mol. Biol. Evol.*, vol. 23, pp. 327–337, 2006.

[113] C. Pal, B. Papp, and L. D. Hurst, “Highly expressed genes in yeast evolve slowly.” *Genetics*, vol. 158, pp. 927–931, 2001.

[114] C. Pal, B. Papp, and M. J. Lercher, "An integrated view of protein evolution." *Nat. Rev. Genet.*, vol. 7, pp. 337–348, 2006.

[115] E. P. Rocha, "The quest for the universals of protein evolution." *Trends Genet.*, vol. 22, pp. 412–416, 2006.

[116] D. P. Wall and et al., "Functional genomic analysis of the rates of protein evolution." *Proc. Natl Acad. Sci. USA*, vol. 102, pp. 5483–5488, 2005.

[117] A. M. Larracuente *et al.*, "Evolution of protein-coding genes in *Drosophila*" *Trends Genet.*, vol. 24, no. 3, pp. 114–123, Mar. 2008.

[118] D. J. Begun *et al.*, "Population Genomics: Whole-Genome Analysis of Polymorphism and Divergence in *Drosophila simulans*" *PLoS Biol.*, vol. 5, no. 11, p. e310, 2007.

[119] A. Stark *et al.*, "Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures" *Nature*, vol. 450, no. 7167, pp. 219–232, 2007.

[120] T. B. Sackton, B. P. Lazzaro, T. Schlenke, J. D. Evans, D. Hultmark, and A. G. Clark, "Dynamic evolution of the innate immune system in *Drosophila*" *Nat. Genet.*, vol. 39, pp. 1461–1468, 2007.

[121] R. Morales-Hojas, C. P. Vieira, M. Reis, and J. Vieira, "Comparative analysis of five immunity-related genes reveals different levels of adaptive evolution in the *virilis* and *melanogaster* groups of *Drosophila*." *Heredity (Edinb).*, vol. 102, no. 6, pp. 573–578, 2009.

[122] M. W. Hahn, M. V. Han, and S. G. Han, "Gene family evolution across 12 *Drosophila* genomes" *PLoS Genet.*, vol. 3, no. 11, pp. 2135–2146, 2007.

[123] T. G. O. Consortium, "Gene ontology: Tool for the unification of biology" *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, 2000.

[124] J. D. Storey, "A direct approach to false discovery rates" *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 64, no. 3, pp. 479–498, 2002.

[125] C. S. McBride and J. R. Arguello, "Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily" *Genetics*, vol. 177, no. 3, pp. 1395–1416, 2007.

[126] F. G. Vieira, A. Sánchez-Gracia, and J. Rozas, "Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution" *Genome Biol.*, vol. 8, no. 11, p. R235, 2007.

[127] A. Gardiner, D. Barker, R. K. Butlin, W. C. Jordan, and M. G. Ritchie, "*Drosophila* chemoreceptor gene evolution: Selection, specialization and genome size" *Mol. Ecol.*, vol. 17, no. 7, pp. 1648–1657, 2008.

[128] C. S. McBride, "Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*." *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 12, pp. 4996–5001, 2007.

[129] N. Tijet, C. Helvig, and R. Feyereisen, "The *cytochrome P450* gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny." *Gene*, vol. 262, no. 1–2, pp. 189–198, Jan. 2001.

[130] H. Ranson, "Evolution of Supergene Families Associated with Insecticide Resistance" *Science (80-.).*, vol. 298, no. 5591, pp. 179–181, 2002.

[131] C. Claudianos *et al.*, "A deficit of detoxification enzymes: Pesticide sensitivity and environmental response in the honeybee" *Insect Mol. Biol.*, vol. 15, no. 5, pp. 615–636, 2006.

[132] B. Lemaitre and J. Hoffmann, "The host defense of *Drosophila melanogaster*." *Annu. Rev. Immunol.*, vol. 25, pp. 697–743, 2007.

[133] T. A. Schlenke and D. J. Begun, "Natural selection drives *Drosophila* immune system evolution." *Genetics*, vol. 164, no. 4, pp. 1471–1480, Aug. 2003.

[134] P. M. Murphy, "Molecular mimicry and the generation of host defense protein diversity." *Cell*, vol. 72, no. 6. United States, pp. 823–826, Mar-1993.

[135] A. L. Hughes and M. Nei, "Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection." *Nature*, vol. 335, no. 6186, pp. 167–170, Sep. 1988.

[136] A. Civetta and R. S. Singh, "High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species" *J. Mol. Evol.*, vol. 41, no. 6, pp. 1085–1095, 1995.

[137] A. Civetta, "Shall we dance or shall we fight? Using DNA sequence data to untangle controversies surrounding sexual selection." *Genome*, vol. 46, no. 6, pp. 925–929, Dec. 2003.

[138] N. L. Clark, J. E. Aagaard, and W. J. Swanson, "Evolution of reproductive proteins from animals and plants." *Reproduction*, vol. 131, no. 1, pp. 11–22, Jan. 2006.

[139] W. Haerty *et al.*, "Evolution in the Fast Lane: Rapidly Evolving Sex-Related Genes in *Drosophila*" *Genetics*, vol. 177, no. 3, pp. 1321–1335, Nov. 2007.

[140] J. Mensch, F. Serra, N. J. Lavagnino, H. Dopazo, and E. Hasson, "Positive selection in nucleoporins challenges constraints on early expressed genes in *Drosophila* development" *Genome Biol. Evol.*, vol. 5, no. 11, pp. 2231–2241, 2013.

[141] C. P. Bird, B. E. Stranger, and E. T. Dermitzakis, "Functional variation and evolution of non-coding DNA." *Curr. Opin. Genet. Dev.*, vol. 16, pp. 559–564, 2006.

[142] P. J. Wittkopp, "Evolution of cis-regulatory sequence and function in *Diptera*" *Heredity (Edinb).*, vol. 97, no. 3, pp. 139–147, 2006.

[143] M. Z. Ludwig, N. H. Patel, and M. Kreitman, "Functional analysis of *eve stripe 2* enhancer evolution in *Drosophila*: rules governing conservation and change." *Development*, vol. 125, no. 5, pp. 949–958, 1998.

[144] N. Bierne and A. Eyre-Walker, “The Genomic Rate of Adaptive Amino Acid Substitution in *Drosophila*” *Mol. Biol. Evol.*, vol. 21, no. 7, pp. 1350–1360, Jul. 2004.

[145] S. A. Sawyer, J. Parsch, Z. Zhang, and D. L. Hartl, “Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*” *Proc. Natl. Acad. Sci.* , vol. 104, no. 16, pp. 6504–6510, Apr. 2007.

[146] S. A. Sawyer, R. J. Kulathinal, C. D. Bustamante, and D. L. Hartl, “Bayesian Analysis Suggests that Most Amino Acid Replacements in *Drosophila* Are Driven by Positive Selection” *J. Mol. Evol.*, vol. 57, no. 1, pp. S154–S164, 2003.

[147] N. G. C. Smith and A. Eyre-Walker, “Adaptive protein evolution in *Drosophila*.” *Nature*, vol. 415, no. 6875, pp. 1022–1024, Feb. 2002.

[148] J. J. Welch, “Estimating the genome-wide rate of adaptive protein evolution in *Drosophila*.” *Genetics*, vol. 173, no. 2, pp. 821–837, Jun. 2006.

[149] J. R. David and P. Capy, “Genetic variation of *Drosophila melanogaster* natural populations.” *Trends Genet.*, vol. 4, pp. 106–111, 1988.

[150] S. Oldham, R. Bohni, H. Stocker, W. Brogiolo, and E. Hafen, “Genetic control of size in *Drosophila*” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 355, no. 1399, pp. 945–952, Jul. 2000.

[151] H. Zhang, J. Liu, C. R. Li, B. Momen, R. A. Kohanski, and L. Pick, “Deletion of *Drosophila insulin*-like peptides causes growth defects and metabolic abnormalities” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 46, pp. 19617–19622, Nov. 2009.

[152] J. A. Walker *et al.*, “Reduced growth of *Drosophila* neurofibromatosis 1 mutants reflects a non-cell-autonomous requirement for *GTPase*-Activating Protein activity in larval neurons.” *Genes Dev.*, vol. 20, no. 23, pp. 3311–3323, Dec. 2006.

[153] J. J. Tong, S. E. Schriner, D. McCleary, B. J. Day, and D. C. Wallace, “Life extension through neurofibromin mitochondrial regulation and antioxidant therapy for neurofibromatosis-1 in *Drosophila melanogaster*.” *Nat. Genet.*, vol. 39, no. 4, pp. 476–485, Apr. 2007.

[154] P. Kapahi, B. M. Zid, T. Harper, D. Koslover, V. Sapin, and S. Benzer, “Regulation of lifespan in *Drosophila* by modulation of genes in the *TOR* signaling pathway.” *Curr. Biol.*, vol. 14, no. 10, pp. 885–890, May 2004.

[155] T. Ikeya, S. Broughton, N. Alic, R. Grandison, and L. Partridge, “The endosymbiont *Wolbachia* increases *insulin*/IGF-like signalling in *Drosophila*.” *Proceedings. Biol. Sci.*, vol. 276, no. 1674, pp. 3799–3807, Nov. 2009.

[156] S. Grönke, D.-F. Clarke, S. Broughton, T. D. Andrews, and L. Partridge, “Molecular Evolution and Functional Characterization of *Drosophila Insulin*-Like Peptides” *PLoS Genet.*, vol. 6, no. 2, p. e1000857, Feb. 2010.

[157] P. D’Avino and C. S. Thummel, “The *Ecdysone* Regulatory Pathway Controls Wing

Morphogenesis and *Integrin* Expression during *Drosophila* Metamorphosis", vol. 220. 2000.

[158] A. F. Simon, C. Shih, A. Mack, and S. Benzer, "Steroid control of longevity in *Drosophila melanogaster*." *Science*, vol. 299, no. 5611, pp. 1407–1410, Feb. 2003.

[159] J. Costas, C. P. Vieira, F. Casares, and J. Vieira, "Genomic characterization of a repetitive motif strongly associated with developmental genes in *Drosophila*" *BMC Genomics*, vol. 4, no. 1, p. 52, 2003.

[160] A. Wodarz and R. Nusse, "Mechanisms of *Wnt* signaling in development." *Annu. Rev. Cell Dev. Biol.*, vol. 14, pp. 59–88, 1998.

[161] C. M. Bergman *et al.*, "Assessing the impact of comparative genomic sequence data on the functional annotation of the *Drosophila* genome" *Genome Biol.*, vol. 3, no. 12, p. research0086.1-86.2, Dec. 2002.

[162] E. M. Zdobnov *et al.*, "Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*." *Science*, vol. 298, no. 5591, pp. 149–159, Oct. 2002.

[163] Y. Yan, H. Wang, H. Chen, A. Lindström-Battle, and R. Jiao, "*Ecdysone* and *Insulin* Signaling Play Essential Roles in Readjusting the Altered Body Size Caused by the *dGPAT4* Mutation in *Drosophila*" *J. Genet. Genomics*, vol. 42, no. 9, pp. 487–494, 2015.

[164] P. Simpson, P. Berreur, and J. Berreur-Bonnenfant, "The initiation of pupariation in *Drosophila*: dependence on growth of the imaginal discs." *J. Embryol. Exp. Morphol.*, vol. 57, pp. 155–165, Jun. 1980.

[165] B. C. Stieper, M. Kupershtok, M. V Driscoll, and A. W. Shingleton, "Imaginal discs regulate developmental timing in *Drosophila melanogaster*." *Dev. Biol.*, vol. 321, no. 1, pp. 18–26, Sep. 2008.

[166] A. Halme, M. Cheng, and I. K. Hariharan, "*Retinoids* regulate a developmental checkpoint for tissue regeneration in *Drosophila*." *Curr. Biol.*, vol. 20, no. 5, pp. 458–463, Mar. 2010.

[167] J. Colombani, D. S. Andersen, and P. Leopold, "Secreted peptide *Dilp8* coordinates *Drosophila* tissue growth with developmental timing." *Science*, vol. 336, no. 6081, pp. 582–585, May 2012.

[168] A. Garelli, A. M. Gontijo, V. Miguela, E. Caparros, and M. Dominguez, "Imaginal discs secrete *insulin*-like peptide 8 to mediate plasticity of growth and maturation." *Science*, vol. 336, no. 6081, pp. 579–582, May 2012.

[169] A. L. Gündner, I. Hahn, O. Sendscheid, H. Aberle, and M. Hoch, "The *PIKE* Homolog *Centaurin gamma* Regulates Developmental Timing in *Drosophila*" *PLoS One*, vol. 9, no. 5, p. e97332, May 2014.

[170] J. Colombani *et al.*, "Antagonistic actions of *ecdysone* and *insulins* determine final size in *Drosophila*." *Science*, vol. 310, no. 5748, pp. 667–670, Oct. 2005.

[171] S. Layalle, N. Arquier, and P. Leopold, "The *TOR* pathway couples nutrition and developmental timing in *Drosophila*." *Dev. Cell*, vol. 15, no. 4, pp. 568–577, Oct. 2008.

[172] J. M. Tennessen and C. S. Thummel, "Coordinating growth and maturation - insights from *Drosophila*." *Curr. Biol.*, vol. 21, no. 18, pp. R750-7, Sep. 2011.

[173] A. W. Shingleton, "Body-size regulation: combining genetics and physiology." *Curr. Biol.*, vol. 15, no. 20, pp. R825-7, Oct. 2005.

[174] C. K. Mirth and A. W. Shingleton, "Integrating body and organ size in *Drosophila*: recent advances and outstanding problems." *Front. Endocrinol. (Lausanne).*, vol. 3, p. 49, 2012.

[175] V. Callier and H. F. Nijhout, "Body size determination in insects: a review and synthesis of size- and brain-dependent and independent mechanisms." *Biol. Rev. Camb. Philos. Soc.*, vol. 88, no. 4, pp. 944–954, Nov. 2013.

[176] D. Reboiro-Jato, M. Reboiro-Jato, F. Fdez-Riverola, C. P. Vieira, N. A. Fonseca, and J. Vieira, "ADOPS--Automatic Detection Of Positively Selected Sites." *J. Integr. Bioinform.*, vol. 9, no. 3, p. 200, Jul. 2012.

[177] O. Folmer, M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek, "DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates." *Mol. Mar. Biol. Biotechnol.*, vol. 3, no. 5, pp. 294–299, Oct. 1994.

[178] F. Ponton, M.-P. Chapuis, M. Pernice, G. A. Sword, and S. J. Simpson, "Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in *Drosophila melanogaster*." *J. Insect Physiol.*, vol. 57, no. 6, pp. 840–850, Jun. 2011.

[179] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." *Methods*, vol. 25, no. 4, pp. 402–408, Dec. 2001.

[180] E. Wittkorn, A. Sarkar, K. Garcia, M. Kango-Singh, and A. Singh, "The *Hippo* pathway effector *Yki* downregulates *Wg* signaling to promote retinal differentiation in the *Drosophila* eye." *Development*, vol. 142, no. 11, pp. 2002–2013, Jun. 2015.

[181] M. Reis, F. B. Valer, C. P. Vieira, and J. Vieira, "*Drosophila Americana* diapausing females show features typical of young flies" *PLoS One*, vol. 10, no. 9, pp. 1–18, 2015.

[182] D. E. Ayer, "Histone deacetylases: transcriptional repression with *SINers* and *NuRDs*." *Trends Cell Biol.*, vol. 9, no. 5, pp. 193–198, May 1999.

[183] A. Grzenda, G. Lomberk, J.-S. Zhang, and R. Urrutia, "*Sin3*: master scaffold and transcriptional corepressor." *Biochim. Biophys. Acta*, vol. 1789, no. 6–8, pp. 443–450, 2009.

[184] A. Swaminathan and L. A. Pile, "Regulation of cell proliferation and wing development by *Drosophila SIN3* and String" *Mech. Dev.*, vol. 127, no. 1–2, pp. 96–106, 2010.

[185] S. M. Cowley *et al.*, "The *mSin3A* chromatin-modifying complex is essential for embryogenesis and T-cell development." *Mol. Cell. Biol.*, vol. 25, no. 16, pp. 6990–7004, Aug. 2005.

[186] J.-H. Dannenberg, G. David, S. Zhong, J. van der Torre, W. H. Wong, and R. A. Depinho, "*mSin3A* corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival." *Genes Dev.*, vol. 19, no. 13, pp. 1581–1595, Jul. 2005.

[187] G. David, K. B. Grandinetti, P. M. Finnerty, N. Simpson, G. C. Chu, and R. A. Depinho, "Specific requirement of the chromatin modifier *mSin3B* in cell cycle exit and cellular differentiation." *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 11, pp. 4168–4172, Mar. 2008.

[188] T. P. Neufeld, A. F. de la Cruz, L. A. Johnston, and B. A. Edgar, "Coordination of growth and cell division in the *Drosophila* wing." *Cell*, vol. 93, no. 7, pp. 1183–1193, Jun. 1998.

[189] G. Pennetta and D. Pauli, "The *Drosophila Sin3* gene encodes a widely distributed transcription factor essential for embryonic viability." *Dev. Genes Evol.*, vol. 208, no. 9, pp. 531–536, Nov. 1998.

[190] L. A. Pile, E. M. Schlag, and D. A. Wassarman, "The *SIN3*/*RPD3* Deacetylase Complex Is Essential for G(2) Phase Cell Cycle Progression and Regulation of *SMRTER* Corepressor Levels" *Mol. Cell. Biol.*, vol. 22, no. 14, pp. 4965–4976, Jul. 2002.

[191] L. A. Pile, P. T. Spellman, R. J. Katzenberger, and D. A. Wassarman, "The *SIN3* deacetylase complex represses genes encoding mitochondrial proteins: implications for the regulation of energy metabolism." *J. Biol. Chem.*, vol. 278, no. 39, pp. 37840–37848, Sep. 2003.

[192] V. Sharma, A. Swaminathan, R. Bao, and L. A. Pile, "*Drosophila SIN3* is required at multiple stages of development." *Dev. Dyn.*, vol. 237, no. 10, pp. 3040–3050, Oct. 2008.

[193] S. Layalle, E. Coessens, A. Ghysen, and C. Dambly-Chaudière, "*Smooth*, a hnRNP encoding gene, controls axonal navigation in *Drosophila*" *Genes to Cells*, vol. 10, no. 2, pp. 119–125, 2005.

[194] D. Paik *et al.*, "Misexpression screen delineates novel genes controlling *Drosophila* lifespan" *Mech. Ageing Dev.*, vol. 133, no. 5, pp. 234–245, 2012.

[195] D. R. Hipfner and S. M. Cohen, "Connecting proliferation and apoptosis in development and disease." *Nat. Rev. Mol. Cell Biol.*, vol. 5, no. 10, pp. 805–815, Oct. 2004.

[196] L. A. Johnston and P. Gallant, "Control of growth and organ size in *Drosophila*." *Bioessays*, vol. 24, no. 1, pp. 54–64, Jan. 2002.

[197] I. Conlon and M. Raff, "Size control in animal development." *Cell*, vol. 96, no. 2, pp. 235–244, Jan. 1999.

[198] R. S. Udan, M. Kango-Singh, R. Nolo, C. Tao, and G. Halder, "*Hippo* promotes proliferation arrest and apoptosis in the *Salvador*/*Warts* pathway." *Nat. Cell Biol.*, vol. 5, no. 10, pp. 914–920, Oct. 2003.

[199] S. Pantalacci, N. Tapon, and P. Leopold, "The *Salvador* partner Hippo promotes apoptosis and cell-cycle exit in Drosophila." *Nat. Cell Biol.*, vol. 5, no. 10, pp. 921–927, Oct. 2003.

[200] S. Wu, J. Huang, J. Dong, and D. Pan, "*hippo* encodes a *Ste-20* family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with *salvador* and *warts*." *Cell*, vol. 114, no. 4, pp. 445–456, Aug. 2003.

[201] K. F. Harvey, C. M. Pfleger, and I. K. Hariharan, "The *Drosophila Mst* ortholog, *hippo*, restricts growth and cell proliferation and promotes apoptosis." *Cell*, vol. 114, no. 4, pp. 457–467, Aug. 2003.

[202] J. Huang, S. Wu, J. Barrera, K. Matthews, and D. Pan, "The *Hippo* signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating *Yorkie*, the *Drosophila* Homolog of *YAP*." *Cell*, vol. 122, no. 3, pp. 421–434, Aug. 2005.

[203] B. A. Edgar, "From cell structure to transcription: *Hippo* forges a new path." *Cell*, vol. 124, no. 2, pp. 267–273, Jan. 2006.

[204] N. Tapon *et al.*, "*salvador* Promotes both cell cycle exit and apoptosis in *Drosophila* and is mutated in human cancer cell lines." *Cell*, vol. 110, no. 4, pp. 467–478, Aug. 2002.

[205] R. W. Justice, O. Zilian, D. F. Woods, M. Noll, and P. J. Bryant, "The *Drosophila* tumor suppressor gene *warts* encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation." *Genes Dev.*, vol. 9, no. 5, pp. 534–546, Mar. 1995.

[206] T. Xu, W. Wang, S. Zhang, R. A. Stewart, and W. Yu, "Identifying tumor suppressors in genetic mosaics: the *Drosophila* lats gene encodes a putative protein kinase." *Development*, vol. 121, no. 4, pp. 1053–1063, Apr. 1995.

[207] M. Kango-Singh *et al.*, "*Shar-pei* mediates cell proliferation arrest during imaginal disc growth in *Drosophila*." *Development*, vol. 129, no. 24, pp. 5719–5730, Dec. 2002.

[208] Z.-C. Lai *et al.*, "Control of cell proliferation and apoptosis by *mob* as tumor suppressor, *mats*." *Cell*, vol. 120, no. 5, pp. 675–685, Mar. 2005.

[209] F. Hamaratoglu *et al.*, "The tumour-suppressor genes *NF2*/*Merlin* and *Expanded* act through *Hippo* signalling to regulate cell proliferation and apoptosis." *Nat. Cell Biol.*, vol. 8, no. 1, pp. 27–36, Jan. 2006.

[210] R. Baumgartner, I. Poernbacher, N. Buser, E. Hafen, and H. Stocker, "The WW domain protein *Kibra* acts upstream of *Hippo* in *Drosophila*." *Dev. Cell*, vol. 18, no. 2, pp. 309–

316, Feb. 2010.

[211] A. Genevet, M. C. Wehr, R. Brain, B. J. Thompson, and N. Tapon, "*Kibra* is a regulator of the *Salvador*/*Warts*/*Hippo* signaling network." *Dev. Cell*, vol. 18, no. 2, pp. 300–308, Feb. 2010.

[212] C.-L. Chen *et al.*, "The apical-basal cell polarity determinant *Crumbs* regulates *Hippo* signaling in *Drosophila*." *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 36, pp. 15810–15815, Sep. 2010.

[213] N. A. Grzeschik, L. M. Parsons, M. L. Allott, K. F. Harvey, and H. E. Richardson, "*Lgl*, *aPKC*, and *Crumbs* regulate the *Salvador*/*Warts*/*Hippo* pathway through two distinct mechanisms." *Curr. Biol.*, vol. 20, no. 7, pp. 573–581, Apr. 2010.

[214] F. C. Bennett and K. F. Harvey, "Fat cadherin modulates organ size in *Drosophila* via the *Salvador*/*Warts*/*Hippo* signaling pathway." *Curr. Biol.*, vol. 16, no. 21, pp. 2101–2110, Nov. 2006.

[215] R. Nolo, C. M. Morrison, C. Tao, X. Zhang, and G. Halder, "The *bantam* MicroRNA Is a Target of the *Hippo* Tumor-Suppressor Pathway" *Curr. Biol.*, vol. 16, no. 19, pp. 1895–1904, 2006.

[216] P. A. Rubenstein, "The functional importance of multiple *actin* isoforms." *Bioessays*, vol. 12, no. 7, pp. 309–315, Jul. 1990.

[217] E. A. Fyrberg, J. W. Mahaffey, B. J. Bond, and N. Davidson, "Transcripts of the six *Drosophila actin* genes accumulate in a stage- and tissue-specific manner" *Cell*, vol. 33, no. 1, pp. 115–123, 1983.

[218] T. L. Lovato, S. M. Meadows, P. W. Baker, J. C. Sparrow, and R. M. Cripps, "Characterization of muscle *actin* genes in *Drosophila virilis* reveals significant molecular complexity in skeletal muscle types" *Insect Mol. Biol.*, vol. 10, no. 4, pp. 333–340, 2001.

VII. Supplementary Data

SUPPLEMENTARY TABLE 1 - Sequence of designed primers that didn't match efficiency levels (90-100%).

Gene	Primer Forward	Primer Reverse
Nurf38	CGAAACGGACTGGAAAATC;	AATGATGGAGGTGGCAAAC
hppy	GCTTATCGGTCTGCTCAATG;	CACAGGTGGCGTATTATTCA
Sin3A	CTCACCCAGACGACCATAC;	TAAGGCGATACACCCACAC
Ih	CAGCAGCAGTCGTCTATCG;	ACTCGTGACAGGCATCGTA
stj	TCGAGCACCGGATGTGATAA;	ACGCATAAAGCCAGTCGAAC
Mmp1	TTCAAGTTGGACGAGGACGA;	CTAGTGGCACCTTTGGAGGA
Pip5K59B	AGAACAGCAGCTCGTCGATA;	TCCGAAACGCTCACAATGTC
Eps-15	GCAGACATCTGGGCTCTTTG;	TAAGTGGGCTTGACCTCCTG
retn	TGATGATCCCAAGCGCAAAG;	ATCTCCTGCCACAGCTTCTT
GLS	CACTCTTATGTGGGCCAGGA;	ACAGCCGTTTGAACCAAGAC

SUPPLEMENTARY TABLE 2 - An analysis of 71 different GO terms lists related with development according to number of matches, matches associated with a chromosome, and by chromosome, 3R, 2R, 3L, 2L, X and 4. CHITEST p values are shown in the last column. (N.A. = not analyzed) The results are presented in total number and relative percentage of genes associated with each of the 6 different chromosomes. The expected number of genes per chromosome is based on information on the number of total coding genes present in each chromosome of *D. melanogaster* reported in "Evolution of genes and genomes on the Drosophila phylogeny" [Artigo1].

GO Term list	GO List number	Matches	Associated with a chromosome	Chromossome 3R (26%)	Chromossome 2R (20%)	Chromossome 3L (20%)	Chromossome 2L (19%)	Chromossome X (14%)	Chromossome 4 (1%)	ChiTEST p value	ChiTEST p value w/Chrx.4
Developmental process	GO:0032502	5506	3694	922 (25%)	734 (20%)	764 (21%)	644 (17%)	594 (16%)	36 (1%)	0,99	0,96
Anatomical structure development	GO:0048856	5165	3491	853 (24%)	709 (20%)	732 (21%)	600 (17%)	563 (16%)	34 (1%)	0,98	0,95
System development	GO:0048731	3142	2416	604 (25%)	485 (20%)	487 (20%)	419 (17%)	396 (16%)	25 (1%)	0,99	0,97
Cellular developmental process	GO:0048869	2944	2485	611 (25%)	481 (19%)	498 (20%)	475 (19%)	400 (16%)	20 (1%)	1,00	0,98
Cell differentiation	GO:0030154	2857	2425	600 (25%)	468 (19%)	487 (20%)	464 (19%)	387 (16%)	19 (1%)	1,00	0,98
Anatomical structure morphogenesis	GO:0009653	2395	1808	452 (25%)	353 (20%)	358 (20%)	323 (18%)	300 (17%)	22 (1%)	0,98	0,95
Nervous system development	GO:0007399	1950	1691	418 (25%)	344 (20%)	343 (20%)	294 (17%)	274 (16%)	18 (1%)	0,99	0,97
Animal Organ development	GO:0048513	1810	1390	351 (25%)	286 (21%)	277 (20%)	241 (17%)	219 (16%)	16 (1%)	0,99	0,96
Neurogenesis	GO:0022008	1661	1483	365 (25%)	302 (20%)	306 (21%)	258 (17%)	239 (16%)	13 (1%)	0,99	0,96
Tissue development	GO:0009888	1398	1172	275 (23%)	215 (18%)	226 (19%)	231 (20%)	215 (18%)	10 (1%)	0,88	0,77
Ephitelium development	GO:0060429	1336	1101	261 (24%)	202 (18%)	206 (19%)	220 (20%)	203 (18%)	9 (1%)	0,90	0,81
Germ cell development	GO:0007281	1262	955	222 (23%)	193 (20%)	168 (18%)	210 (22%)	158 (17%)	4 (0%)	0,75	0,80
Organ morphogenesis	GO:0009887	1114	889	220 (25%)	178 (20%)	176 (20%)	151 (17%)	153 (17%)	11 (1%)	0,97	0,93
Oogenesis	GO:0048477	1051	841	189 (22%)	174 (21%)	143 (17%)	183 (22%)	149 (18%)	3 (0%)	0,59	0,60
Regulation of developmental process	GO:0050793	899	757	186 (25%)	139 (18%)	147 (19%)	133 (18%)	141 (19%)	11 (1%)	0,83	0,71
Tissue morphogenesis	GO:0048729	864	709	164 (23%)	137 (19%)	142 (20%)	125 (18%)	134 (19%)	7 (1%)	0,82	0,69
Post-embryonic development	GO:0009791	823	755	180 (24%)	142 (19%)	157 (21%)	120 (16%)	149 (20%)	7 (1%)	0,65	0,51
Instar larval or pupal development	GO:0002165	750	683	155 (23%)	129 (19%)	144 (21%)	105 (15%)	143 (21%)	7 (1%)	0,44	0,31
Sensory organ development	GO:0007423	749	498	135 (27%)	102 (20%)	90 (18%)	93 (19%)	72 (14%)	6 (1%)	1,00	0,99
Embryo development	GO:0009790	728	535	141 (26%)	107 (20%)	93 (17%)	96 (18%)	93 (17%)	5 (1%)	0,95	0,89

Imaginal disc development	GO:0007444	724	670	158 (24%)	122 (18%)	129 (19%)	126 (19%)	126 (19%)	9 (1%)	0,82	0,70
Post-embryonic animal morphogenesis	GO:0009886	655	589	132 (22%)	115 (20%)	121 (21%)	95 (16%)	119 (20%)	7 (1%)	0,59	0,45
Metamorphosis	GO:0007552	649	601	132 (22%)	116 (19%)	130 (22%)	99 (16%)	117 (19%)	7 (1%)	0,68	0,54
Instar larval or pupal morphogenesis	GO:0048707	643	577	128 (22%)	113 (20%)	119 (21%)	94 (16%)	116 (20%)	7 (1%)	0,59	0,45
Eye development	GO:0001654	559	380	92 (24%)	81 (21%)	68 (18%)	74 (19%)	59 (16%)	6 (2%)	0,89	0,95
Regulation of multicellular organismal development	GO:2000026	544	502	132 (26%)	96 (19%)	100 (20%)	85 (17%)	84 (17%)	5 (1%)	0,97	0,92
Compound eye development	GO:0048749	535	363	88 (24%)	79 (22%)	66 (18%)	71 (20%)	54 (15%)	5 (1%)	0,98	0,95
Wing disc development	GO:0035220	522	484	111 (23%)	95 (20%)	94 (19%)	88 (18%)	90 (19%)	6 (1%)	0,82	0,69
Imagina disc morphogenesis	GO:0007560	516	477	97 (20%)	90 (19%)	94 (20%)	85 (18%)	104 (22%)	7 (1%)	0,30	0,19
Appendage development	GO:0048736	467	428	91 (21%)	80 (19%)	89 (21%)	70 (16%)	93 (22%)	5 (1%)	0,30	0,19
Imaginal disc derived appendage development	GO:0048737	462	423	90 (21%)	79 (19%)	87 (21%)	69 (16%)	93 (22%)	5 (1%)	0,30	0,19
Appendage morphogenesis	GO:0035107	458	419	87 (21%)	79 (19%)	88 (21%)	68 (16%)	92 (22%)	5 (1%)	0,30	0,19
Imaginal disc derived appendage morphogenesis	GO:0035114	455	416	87 (21%)	78 (19%)	87 (21%)	67 (16%)	92 (22%)	5 (1%)	0,30	0,19
Eye morphogenesis	GO:0048592	450	301	71 (24%)	67 (22%)	53 (18%)	61 (20%)	45 (15%)	4 (1%)	0,99	0,97
Developmental growth	GO:0048589	423	376	90 (24%)	68 (18%)	86 (23%)	55 (15%)	73 (19%)	4 (1%)	0,63	0,49
Wing disc morphogenesis	GO:0007472	403	368	75 (20%)	75 (20%)	72 (20%)	61 (17%)	80 (22%)	5 (1%)	0,29	0,19
Ovarian follicle cell development	GO:0030707	386	327	65 (20%)	60 (18%)	48 (15%)	80 (24%)	72 (22%)	2 (1%)	0,12	0,07
Wnt signaling pathway	GO:0016055	356	145	37 (26%)	26 (18%)	35 (24%)	20 (14%)	23 (16%)	4 (3%)	0,61	0,63
Regulation of growth	GO:0040008	354	328	82 (25%)	64 (20%)	72 (22%)	40 (12%)	64 (20%)	6 (2%)	0,27	0,25
Regulation of nervous system development	GO:0051960	337	311	79 (25%)	64 (21%)	65 (21%)	45 (14%)	54 (17%)	4 (1%)	0,84	0,72
Open tracheal system development	GO:0007424	326	238	55 (23%)	49 (21%)	46 (19%)	48 (20%)	37 (16%)	3 (1%)	0,98	0,94
Regulation of anatomical structure morphogenesis	GO:0022603	318	302	69 (23%)	48 (16%)	50 (17%)	62 (21%)	68 (23%)	5 (2%)	0,13	0,11
Central nervous system development	GO:0007417	302	279	60 (22%)	60 (22%)	57 (20%)	48 (17%)	48 (17%)	6 (2%)	0,75	0,80

Embryonic morphogenesis	GO:0048598	286	242	65 (27%)	45 (19%)	39 (16%)	47 (19%)	45 (19%)	1 (0%)	0,60	0,61
Cuticle development	GO:0042335	275	270	74 (27%)	52 (19%)	80 (30%)	33 (12%)	29 (11%)	2 (1%)	0,14	0,08
Embryonic development via the synctial blastoderm	GO:0001700	268	187	45 (24%)	40 (21%)	32 (17%)	34 (18%)	35 (19%)	1 (1%)	0,78	0,65
Regulation of developmental growth	GO:0048638	247	245	61 (25%)	46 (19%)	56 (23%)	32 (13%)	45 (18%)	5 (2%)	0,47	0,47
Gland development	GO:0048732	245	241	67 (28%)	52 (22%)	52 (22%)	39 (16%)	30 (12%)	1 (0%)	0,80	0,86
Anterior/Posterior pattern specification	GO:0009952	220	174	48 (28%)	44 (25%)	16 (9%)	37 (21%)	28 (16%)	1 (1%)	0,16	0,09
Negative regulation of developmental process	GO:0051093	219	200	59 (30%)	32 (16%)	38 (19%)	40 (20%)	30 (15%)	1 (1%)	0,90	0,81
Salivary gland development	GO:0007431	188	188	51 (27%)	42 (22%)	44 (23%)	26 (14%)	24 (13%)	1 (1%)	0,84	0,72
Exocrine system development	GO:0035272	188	188	51 (27%)	42 (22%)	44 (23%)	26 (14%)	24 (13%)	1 (1%)	0,84	0,72
Oocyte development	GO:0048599	188	146	42 (29%)	37 (25%)	19 (13%)	28 (19%)	20 (14%)	0 (0%)	0,41	0,40
Determination of adult lifespan	GO:0008340	172	169	42 (25%)	33 (20%)	35 (21%)	28 (17%)	31 (18%)	0 (0%)	0,79	0,84
Muscle organ development	GO:0007517	171	141	34 (24%)	38 (27%)	24 (17%)	22 (16%)	22 (16%)	1 (1%)	0,58	0,43
Brain development	GO:0007420	150	135	31 (23%)	28 (21%)	21 (16%)	22 (16%)	27 (20%)	6 (4%)	0,02	0,37
Developmental pigmentation	GO:0048066	131	118	37 (31%)	13 (11%)	23 (19%)	20 (17%)	24 (20%)	1 (1%)	0,17	0,10
Heart development	GO:0007507	111	105	28 (27%)	22 (21%)	17 (16%)	24 (23%)	12 (11%)	2 (2%)	0,64	0,67
Bristle/Chaeta development	GO:0022416	104	101	35 (35%)	11 (11%)	19 (19%)	19 (19%)	17 (17%)	0 (0%)	0,11	0,10
Imaginal disc pattern formation	GO:0007447	104	103	29 (28%)	19 (18%)	20 (19%)	23 (22%)	11 (11%)	1 (1%)	0,91	0,82
Leg morphogenesis	GO:0007478	87	85	19 (22%)	14 (16%)	21 (25%)	12 (14%)	19 (22%)	0 (0%)	0,09	0,07
Molting cycle process	GO:0022404	72	63	18 (29%)	9 (14%)	17 (27%)	9 (14%)	9 (14%)	1 (2%)	0,23	0,21
Muscle cell development	GO:0055001	61	60	11 (18%)	12 (20%)	15 (25%)	8 (13%)	13 (22%)	1 (2%)	0,05	**0,04**
Response to ecdysone	GO:0035075	54	44	6 (14%)	7 (14%)	14 (32%)	9 (20%)	8 (18%)	0 (0%)	0,01	**0,00**
Imaginal disc growth	GO:0007446	53	47	17 (36%)	5 (11%)	9 (19%)	10 (21%)	5 (11%)	1 (2%)	0,08	0,07
Skeletal muscle development	GO:0060538	53	52	11 (21%)	11 (21%)	10 (19%)	10 (19%)	10 (19%)	0 (0%)	0,57	0,58

Response to insulin	GO:0032868	46	46	14 (30%)	3 (7%)	16 (35%)	7 (15%)	6 (13%)	0 (0%)	0,00	**0,00**
Striated muscle development	GO:0014706	32	28	7 (25%)	6 (21%)	4 (14%)	5 (18%)	6 (21%)	0 (0%)	0,27	0,24
Ectoderm development	GO:0007398	29	29	8 (28%)	3 (10%)	2 (7%)	8 (28%)	8 (28%)	0 (0%)	0,00	**0,00**
Muscle fiber development	GO:0048747	13	12	2 (17%)	5 (42%)	3 (25%)	1 (8%)	1 (8%)	0 (0%)	0,00	**0,00**
Limb development	GO:0060173	3	3	1 (33%)	0 (0%)	1 (33%)	1 (33%)	0 (0%)	0 (0%)	N.A.	N.A.

Supplementary Table 3 | RT-PCR CT values for *Sin3A*

For each one of the 12 different conditions, CT values for 3 different biological replicates (Replicate from #A-#C) are shown and for each, we made 2 technical replicates (CT.1 and CT.2), the mean value of this replicates is shown. The comparison of cycle value for candidate gene with reference gene is shown in ΔCT.

Strain	Conditions	Replicate	Gene	CT.1	CT.2	Mean	Gene	CT.1	CT.2	Mean	ΔCT
W11	15x4	A	*Sin3A*	25,42	25,29	25,36	*RPL32*	24,11	24,29	24,20	1,16
W11	15x4	B	*Sin3A*	25,11	24,98	25,05	*RPL32*	26,38	26,99	26,69	-1,64
W11	15x4	C	*Sin3A*	24,36	24,53	24,45	*RPL32*	25,49	26,02	25,76	-1,31
W11	15x8	A	*Sin3A*	26,43	26,72	26,58	*RPL32*	25,30	25,30	25,30	1,28
W11	15x8	B	*Sin3A*	25,89	26,37	26,13	*RPL32*	25,30	25,44	25,37	0,76
W11	15x8	C	*Sin3A*	25,18	25,40	25,29	*RPL32*	25,45	25,73	25,59	-0,30
W11	16x4	A	*Sin3A*	26,53	26,25	26,39	*RPL32*	25,24	24,97	25,11	1,29
W11	16x4	B	*Sin3A*	26,48	26,75	26,62	*RPL32*	24,64	24,52	24,58	2,04
W11	16x4	C	*Sin3A*	24,47	24,68	24,58	*RPL32*	25,27	25,20	25,24	-0,66
W11	16x8	A	*Sin3A*	28,02	27,80	27,91	*RPL32*	26,25	26,57	26,41	1,50
W11	16x8	B	*Sin3A*	25,02	24,85	24,94	*RPL32*	25,46	25,89	25,68	-0,74
W11	16x8	C	*Sin3A*	24,16	24,18	24,17	*RPL32*	25,09	25,06	25,08	-0,90
SF12	15x4	A	*Sin3A*	27,21	27,18	27,20	*RPL32*	27,06	27,30	27,18	0,02
SF12	15x4	B	*Sin3A*	26,14	26,25	26,20	*RPL32*	27,57	27,93	27,75	-1,56
SF12	15x4	C	*Sin3A*	26,27	26,52	26,40	*RPL32*	26,43	26,46	26,45	-0,05
SF12	15x8	A	*Sin3A*	27,18	27,08	27,13	*RPL32*	27,41	27,50	27,46	-0,32
SF12	15x8	B	*Sin3A*	26,36	26,63	26,50	*RPL32*	26,84	27,11	26,98	-0,48
SF12	15x8	C	*Sin3A*	27,02	27,20	27,11	*RPL32*	27,35	27,30	27,33	-0,22
SF12	16x4	A	*Sin3A*	26,07	26,11	26,09	*RPL32*	27,20	27,21	27,21	-1,12
SF12	16x4	B	*Sin3A*	26,26	26,39	26,33	*RPL32*	26,24	26,37	26,31	0,02
SF12	16x4	C	*Sin3A*	25,41	25,89	25,65	*RPL32*	26,62	26,86	26,74	-1,09
SF12	16x8	A	*Sin3A*	26,54	26,68	26,61	*RPL32*	26,74	26,87	26,81	-0,20
SF12	16x8	B	*Sin3A*	27,74	27,79	27,77	*RPL32*	27,71	28,26	27,99	-0,22
SF12	16x8	C	*Sin3A*	25,61	25,86	25,74	*RPL32*	26,20	26,36	26,28	-0,55
NOVA00	16x4	A	*Sin3A*	26,25	26,16	26,21	*RPL32*	27,26	27,39	27,33	-1,12
NOVA00	16x4	B	*Sin3A*	26,60	26,65	26,63	*RPL32*	26,80	26,88	26,84	-0,22
NOVA00	16x4	C	*Sin3A*	26,09	26,48	26,29	*RPL32*	27,91	28,05	27,98	-1,70
NOVA00	16x8	A	*Sin3A*	26,72	26,43	26,58	*RPL32*	26,58	26,74	26,66	-0,08
NOVA00	16x8	B	*Sin3A*	26,61	26,84	26,73	*RPL32*	27,71	28,25	27,98	-1,26
NOVA00	16x8	C	*Sin3A*	25,33	25,50	25,42	*RPL32*	26,98	26,97	26,98	-1,56
NOVA00	17x4	A	*Sin3A*	27,24	27,35	27,30	*RPL32*	28,32	28,91	28,62	-1,32
NOVA00	17x4	B	*Sin3A*	26,50	26,75	26,63	*RPL32*	27,36	27,27	27,32	-0,69
NOVA00	17x4	C	*Sin3A*	25,72	25,97	25,85	*RPL32*	27,00	27,33	27,17	-1,32
NOVA00	17x8	A	*Sin3A*	26,83	27,05	26,94	*RPL32*	28,32	28,30	28,31	-1,37
NOVA00	17x8	B	*Sin3A*	26,64	27,09	26,87	*RPL32*	28,58	29,04	28,81	-1,95
NOVA00	17x8	C	*Sin3A*	27,39	27,59	27,49	*RPL32*	28,92	29,01	28,97	-1,48

Supplementary Table 4 | RT-PCR CT values for *CG3209*

For each one of the 12 different conditions, CT values for 3 different biological replicates (Replicate from #A-#C) are shown and for each, we made 2 technical replicates (CT.1 and CT.2), the mean value of this replicates is shown. The comparison of cycle value for candidate gene with reference gene is shown in ΔCT.

Strain	Conditions	Replicate	Gene	CT.1	CT.2	Mean	Gene	CT.1	CT.2	Mean	ΔCT
W11	15x4	A	*CG3209*	24,05	24,09	24,07	*RPL32*	24,11	24,29	24,20	-0,13
W11	15x4	B	*CG3209*	23,83	24,10	23,97	*RPL32*	26,38	26,99	26,69	-2,72
W11	15x4	C	*CG3209*	23,19	23,31	23,25	*RPL32*	25,49	26,02	25,76	-2,51
W11	15x8	A	*CG3209*	24,83	24,90	24,87	*RPL32*	25,30	25,30	25,30	-0,44
W11	15x8	B	*CG3209*	24,21	24,52	24,37	*RPL32*	25,30	25,44	25,37	-1,01
W11	15x8	C	*CG3209*	23,49	24,99	24,24	*RPL32*	25,45	25,73	25,59	-1,35
W11	16x4	A	*CG3209*	24,28	24,14	24,21	*RPL32*	25,24	24,97	25,11	-0,89
W11	16x4	B	*CG3209*	24,74	24,85	24,80	*RPL32*	24,64	24,52	24,58	0,22
W11	16x4	C	*CG3209*	23,13	22,94	23,04	*RPL32*	25,27	25,20	25,24	-2,20
W11	16x8	A	*CG3209*	26,21	26,26	26,24	*RPL32*	26,14	26,22	26,18	0,05
W11	16x8	B	*CG3209*	22,68	23,28	22,98	*RPL32*	25,46	25,89	25,68	-2,70
W11	16x8	C	*CG3209*	22,12	22,42	22,27	*RPL32*	25,09	25,06	25,08	-2,81
SF12	15x4	A	*CG3209*	25,65	25,42	25,54	*RPL32*	26,64	27,07	26,86	-1,32
SF12	15x4	B	*CG3209*	24,86	25,26	25,06	*RPL32*	28,30	27,89	28,10	-3,04
SF12	15x4	C	*CG3209*	24,38	24,90	24,64	*RPL32*	26,27	26,58	26,43	-1,79
SF12	15x8	A	*CG3209*	26,08	25,18	25,63	*RPL32*	26,69	26,88	26,79	-1,16
SF12	15x8	B	*CG3209*	24,18	24,34	24,26	*RPL32*	25,78	26,38	26,08	-1,82
SF12	15x8	C	*CG3209*	24,68	25,15	24,92	*RPL32*	26,10	26,62	26,36	-1,45
SF12	16x4	A	*CG3209*	24,61	24,32	24,47	*RPL32*	27,46	27,42	27,44	-2,98
SF12	16x4	B	*CG3209*	24,20	24,37	24,29	*RPL32*	26,18	26,81	26,50	-2,21
SF12	16x4	C	*CG3209*	24,10	24,10	24,10	*RPL32*	26,24	26,35	26,30	-2,20
SF12	16x8	A	*CG3209*	24,61	24,67	24,64	*RPL32*	26,36	26,40	26,38	-1,74
SF12	16x8	B	*CG3209*	25,58	26,02	25,80	*RPL32*	27,34	27,34	27,34	-1,54
SF12	16x8	C	*CG3209*	23,37	23,58	23,48	*RPL32*	25,61	25,81	25,71	-2,24
NOVA00	16x4	A	*CG3209*	25,13	25,24	25,19	*RPL32*	27,78	27,50	27,64	-2,46
NOVA00	16x4	B	*CG3209*	24,35	24,47	24,41	*RPL32*	26,80	26,88	26,84	-2,43
NOVA00	16x4	C	*CG3209*	24,84	25,08	24,96	*RPL32*	28,09	27,97	28,03	-3,07
NOVA00	16x8	A	*CG3209*	24,58	24,33	24,46	*RPL32*	26,72	26,90	26,81	-2,36
NOVA00	16x8	B	*CG3209*	24,60	24,86	24,73	*RPL32*	27,57	28,00	27,79	-3,06
NOVA00	16x8	C	*CG3209*	24,29	24,54	24,42	*RPL32*	26,74	27,00	26,87	-2,46
NOVA00	17x4	A	*CG3209*	25,89	25,71	25,80	*RPL32*	28,22	28,45	28,34	-2,54
NOVA00	17x4	B	*CG3209*	24,06	24,24	24,15	*RPL32*	26,59	26,88	26,74	-2,59
NOVA00	17x4	C	*CG3209*	23,88	23,88	23,88	*RPL32*	27,16	27,44	27,30	-3,42
NOVA00	17x8	A	*CG3209*	25,08	25,22	25,15	*RPL32*	28,11	28,21	28,16	-3,01
NOVA00	17x8	B	*CG3209*	25,98	25,50	25,74	*RPL32*	28,57	28,51	28,54	-2,80
NOVA00	17x8	C	*CG3209*	25,06	25,46	25,26	*RPL32*	28,73	28,58	28,66	-3,40

Supplementary Table 5 | RT-PCR CT values for *sm*

For each one of the 12 different conditions, CT values for 3 different biological replicates (Replicate from #A-#C) are shown and for each, we made 2 technical replicates (CT.1 and CT.2), the mean value of this replicates is shown. The comparison of cycle value for candidate gene with reference gene is shown in ΔCT.

Strain	Conditions	Replicate	Gene	CT.1	CT.2	Mean	Gene	CT.1	CT.2	Mean	ΔCT
W11	15x4	A	*sm*	22,39	22,03	22,21	*RPL32*	24,11	24,29	24,20	-1,99
W11	15x4	B	*sm*	21,64	21,98	21,81	*RPL32*	26,38	26,99	26,69	-4,88
W11	15x4	C	*sm*	21,09	21,41	21,25	*RPL32*	25,49	26,02	25,76	-4,51
W11	15x8	A	*sm*	22,69	22,36	22,53	*RPL32*	25,30	25,30	25,30	-2,78
W11	15x8	B	*sm*	23,42	23,46	23,44	*RPL32*	25,30	25,44	25,37	-1,93
W11	15x8	C	*sm*	22,50	23,11	22,81	*RPL32*	25,45	25,73	25,59	-2,79
W11	16x4	A	*sm*	22,43	22,33	22,38	*RPL32*	25,24	24,97	25,11	-2,73
W11	16x4	B	*sm*	22,59	22,83	22,71	*RPL32*	24,64	24,52	24,58	-1,87
W11	16x4	C	*sm*	21,13	21,24	21,19	*RPL32*	25,27	25,20	25,24	-4,05
W11	16x8	A	*sm*	23,22	23,02	23,12	*RPL32*	25,14	25,21	25,18	-2,06
W11	16x8	B	*sm*	21,20	21,31	21,26	*RPL32*	25,46	25,89	25,68	-4,42
W11	16x8	C	*sm*	20,56	20,69	20,63	*RPL32*	25,09	25,06	25,08	-4,45
SF12	15x4	A	*sm*	22,44	22,56	22,50	*RPL32*	25,48	25,81	25,65	-3,15
SF12	15x4	B	*sm*	22,06	22,52	22,29	*RPL32*	27,13	27,45	27,29	-5,00
SF12	15x4	C	*sm*	22,18	22,30	22,24	*RPL32*	25,94	25,96	25,95	-3,71
SF12	15x8	A	*sm*	23,05	22,43	22,74	*RPL32*	26,86	26,88	26,87	-4,13
SF12	15x8	B	*sm*	22,17	22,20	22,19	*RPL32*	25,28	25,83	25,56	-3,37
SF12	15x8	C	*sm*	22,14	22,20	22,17	*RPL32*	25,95	26,10	26,03	-3,86
SF12	16x4	A	*sm*	21,57	21,35	21,46	*RPL32*	26,18	26,18	26,18	-4,72
SF12	16x4	B	*sm*	21,00	21,06	21,03	*RPL32*	25,38	25,35	25,37	-4,34
SF12	16x4	C	*sm*	20,85	20,65	20,75	*RPL32*	25,74	25,32	25,53	-4,78
SF12	16x8	A	*sm*	22,16	21,73	21,95	*RPL32*	25,85	25,78	25,82	-3,87
SF12	16x8	B	*sm*	22,61	22,99	22,80	*RPL32*	26,65	27,20	26,93	-4,13
SF12	16x8	C	*sm*	19,98	20,38	20,18	*RPL32*	25,16	24,66	24,91	-4,73
NOVA00	16x4	A	*sm*	22,39	22,22	22,31	*RPL32*	27,58	27,63	27,61	-5,30
NOVA00	16x4	B	*sm*	22,13	22,36	22,25	*RPL32*	29,08	29,21	29,15	-6,90
NOVA00	16x4	C	*sm*	21,46	22,08	21,77	*RPL32*	28,20	28,11	28,16	-6,39
NOVA00	16x8	A	*sm*	22,35	21,64	22,00	*RPL32*	26,37	27,02	26,70	-4,70
NOVA00	16x8	B	*sm*	21,54	21,85	21,70	*RPL32*	27,91	28,04	27,98	-6,28
NOVA00	16x8	C	*sm*	22,06	22,31	22,19	*RPL32*	27,29	26,50	26,90	-4,71
NOVA00	17x4	A	*sm*	22,63	22,69	22,66	*RPL32*	28,40	28,54	28,47	-5,81
NOVA00	17x4	B	*sm*	22,49	22,41	22,45	*RPL32*	27,47	27,32	27,40	-4,95
NOVA00	17x4	C	*sm*	20,92	21,04	20,98	*RPL32*	26,79	27,57	27,18	-6,20
NOVA00	17x8	A	*sm*	21,40	21,44	21,42	*RPL32*	28,04	28,38	28,21	-6,79
NOVA00	17x8	B	*sm*	21,48	21,73	21,61	*RPL32*	28,35	28,77	28,56	-6,96
NOVA00	17x8	C	*sm*	22,57	22,66	22,62	*RPL32*	28,73	28,86	28,80	-6,18

Supplementary Table 6 | RT-PCR CT values for *yki*

For each one of the 12 different conditions, CT values for 3 different biological replicates (Replicate from #A-#C) are shown and for each, we made 2 technical replicates (CT.1 and CT.2), the mean value of this replicates is shown. The comparison of cycle value for candidate gene with reference gene is shown in ΔCT.

Strain	Conditions	Replicate	Gene	CT.1	CT.2	Mean	Gene	CT.1	CT.2	Mean	ΔCT
W11	15x4	A	*yki*	24,57	24,73	24,65	*RPL32*	24,39	25,09	24,74	-0,09
W11	15x4	B	*yki*	24,73	24,73	24,73	*RPL32*	27,57	27,71	27,64	-2,91
W11	15x4	C	*yki*	23,99	24,54	24,27	*RPL32*	26,72	27,02	26,87	-2,61
W11	15x8	A	*yki*	25,75	25,05	25,40	*RPL32*	26,89	27,35	27,12	-1,72
W11	15x8	B	*yki*	25,18	25,58	25,38	*RPL32*	25,23	25,40	25,32	0,07
W11	15x8	C	*yki*	24,07	24,20	24,14	*RPL32*	25,18	25,12	25,15	-1,02
W11	16x4	A	*yki*	24,01	24,00	24,01	*RPL32*	25,03	25,42	25,23	-1,22
W11	16x4	B	*yki*	25,23	25,30	25,27	*RPL32*	25,51	25,57	25,54	-0,27
W11	16x4	C	*yki*	23,73	23,88	23,81	*RPL32*	26,69	26,35	26,52	-2,72
W11	16x8	A	*yki*	23,98	24,29	24,14	*RPL32*	28,78	28,62	28,70	-4,57
W11	16x8	B	*yki*	23,94	24,17	24,06	*RPL32*	25,72	25,85	25,79	-1,73
W11	16x8	C	*yki*	23,11	23,33	23,22	*RPL32*	25,00	25,28	25,14	-1,92
SF12	15x4	A	*yki*	27,42	27,24	27,33	*RPL32*	28,23	28,48	28,36	-1,03
SF12	15x4	B	*yki*	26,58	26,75	26,67	*RPL32*	29,09	29,17	29,13	-2,47
SF12	15x4	C	*yki*	26,49	26,97	26,73	*RPL32*	28,35	28,05	28,20	-1,47
SF12	15x8	A	*yki*	27,59	27,31	27,45	*RPL32*	28,61	28,89	28,75	-1,30
SF12	15x8	B	*yki*	26,34	26,66	26,50	*RPL32*	27,72	27,85	27,79	-1,29
SF12	15x8	C	*yki*	26,44	26,95	26,70	*RPL32*	28,31	28,33	28,32	-1,63
SF12	16x4	A	*yki*	26,36	26,47	26,42	*RPL32*	28,92	29,30	29,11	-2,70
SF12	16x4	B	*yki*	25,92	26,09	26,01	*RPL32*	27,77	27,68	27,73	-1,72
SF12	16x4	C	*yki*	25,46	25,57	25,52	*RPL32*	27,63	28,23	27,93	-2,42
SF12	16x8	A	*yki*	26,74	26,71	26,73	*RPL32*	28,38	28,49	28,44	-1,71
SF12	16x8	B	*yki*	28,58	28,81	28,70	*RPL32*	29,51	29,89	29,70	-1,01
SF12	16x8	C	*yki*	24,77	25,23	25,00	*RPL32*	27,46	27,54	27,50	-2,50
NOVA00	16x4	A	*yki*	25,36	25,21	25,29	*RPL32*	27,33	27,77	27,55	-2,27
NOVA00	16x4	B	*yki*	26,10	26,13	26,12	*RPL32*	27,78	28,36	28,07	-1,96
NOVA00	16x4	C	*yki*	25,79	26,29	26,04	*RPL32*	27,97	28,12	28,05	-2,01
NOVA00	16x8	A	*yki*	24,48	23,95	24,22	*RPL32*	26,84	27,10	26,97	-2,76
NOVA00	16x8	B	*yki*	24,68	24,75	24,72	*RPL32*	27,66	27,94	27,80	-3,09
NOVA00	16x8	C	*yki*	24,87	24,79	24,83	*RPL32*	27,04	27,12	27,08	-2,25
NOVA00	17x4	A	*yki*	25,43	24,17	24,80	*RPL32*	28,31	28,44	28,38	-3,58
NOVA00	17x4	B	*yki*	24,68	24,98	24,83	*RPL32*	26,69	26,63	26,66	-1,83
NOVA00	17x4	C	*yki*	23,05	24,15	23,60	*RPL32*	26,68	27,05	26,87	-3,27
NOVA00	17x8	A	*yki*	24,45	24,66	24,56	*RPL32*	27,60	27,47	27,54	-2,98
NOVA00	17x8	B	*yki*	25,05	25,27	25,16	*RPL32*	28,57	28,69	28,63	-3,47
NOVA00	17x8	C	*yki*	25,33	25,45	25,39	*RPL32*	28,70	28,70	28,70	-3,31

Supplementary Table 7 | RT-PCR CT values for *hpo*

For each one of the 12 different conditions, CT values for 3 different biological replicates (Replicate from #A-#C) are shown and for each, we made 2 technical replicates (CT.1 and CT.2), the mean value of this replicates is shown. The comparison of cycle value for candidate gene with reference gene is shown in ΔCT.

Strain	Conditions	Replicate	Gene	CT.1	CT.2	Mean	Gene	CT.1	CT.2	Mean	ΔCT
W11	15x4	A	*hpo*	26,25	25,99	26,12	*RPL32*	28,03	28,40	28,22	-2,10
W11	15x4	B	*hpo*	26,02	26,27	26,15	*RPL32*	29,37	29,12	29,25	-3,10
W11	15x4	C	*hpo*	25,93	26,11	26,02	*RPL32*	29,06	28,97	29,02	-3,00
W11	15x8	A	*hpo*	25,97	25,53	25,75	*RPL32*	28,24	28,59	28,42	-2,67
W11	15x8	B	*hpo*	25,54	25,63	25,59	*RPL32*	28,49	28,63	28,56	-2,98
W11	15x8	C	*hpo*	25,38	25,83	25,61	*RPL32*	28,50	28,32	28,41	-2,81
W11	16x4	A	*hpo*	26,96	26,90	26,93	*RPL32*	29,35	29,42	29,39	-2,46
W11	16x4	B	*hpo*	27,01	27,02	27,02	*RPL32*	29,49	29,58	29,54	-2,52
W11	16x4	C	*hpo*	26,34	26,51	26,43	*RPL32*	28,97	29,43	29,20	-2,78
W11	16x8	A	*hpo*	26,28	26,13	26,21	*RPL32*	29,47	29,45	29,46	-3,26
W11	16x8	B	*hpo*	25,19	25,37	25,28	*RPL32*	28,33	28,56	28,45	-3,17
W11	16x8	C	*hpo*	25,18	25,46	25,32	*RPL32*	28,07	28,24	28,16	-2,84
SF12	15x4	A	*hpo*	27,94	27,62	27,78	*RPL32*	27,06	27,30	27,18	0,60
SF12	15x4	B	*hpo*	26,75	27,18	26,97	*RPL32*	28,23	27,87	28,05	-1,09
SF12	15x4	C	*hpo*	27,11	27,17	27,14	*RPL32*	27,30	27,28	27,29	-0,15
SF12	15x8	A	*hpo*	28,27	28,06	28,17	*RPL32*	27,80	28,20	28,00	0,16
SF12	15x8	B	*hpo*	26,86	26,83	26,85	*RPL32*	26,61	26,58	26,60	0,25
SF12	15x8	C	*hpo*	27,52	27,61	27,57	*RPL32*	27,29	27,71	27,50	0,06
SF12	16x4	A	*hpo*	26,98	26,84	26,91	*RPL32*	28,03	28,12	28,08	-1,17
SF12	16x4	B	*hpo*	26,75	27,02	26,89	*RPL32*	26,82	26,58	26,70	0,18
SF12	16x4	C	*hpo*	26,32	26,35	26,34	*RPL32*	26,78	27,49	27,14	-0,80
SF12	16x8	A	*hpo*	27,71	27,62	27,67	*RPL32*	27,46	27,53	27,50	0,17
SF12	16x8	B	*hpo*	28,66	28,51	28,59	*RPL32*	28,71	28,09	28,40	0,19
SF12	16x8	C	*hpo*	26,01	26,36	26,19	*RPL32*	26,63	26,83	26,73	-0,54
NOVA00	16x4	A	*hpo*	25,63	25,71	25,67	*RPL32*	27,15	27,24	27,20	-1,53
NOVA00	16x4	B	*hpo*	26,32	26,89	26,61	*RPL32*	27,32	28,22	27,77	-1,17
NOVA00	16x4	C	*hpo*	25,30	26,08	25,69	*RPL32*	27,40	27,03	27,22	-1,53
NOVA00	16x8	A	*hpo*	26,12	25,66	25,89	*RPL32*	26,50	26,74	26,62	-0,73
NOVA00	16x8	B	*hpo*	25,56	25,35	25,46	*RPL32*	26,91	27,32	27,12	-1,66
NOVA00	16x8	C	*hpo*	25,75	26,32	26,04	*RPL32*	26,38	26,47	26,43	-0,39
NOVA00	17x4	A	*hpo*	26,69	27,10	26,90	*RPL32*	28,13	28,33	28,23	-1,33
NOVA00	17x4	B	*hpo*	25,09	25,10	25,10	*RPL32*	26,11	26,39	26,25	-1,16
NOVA00	17x4	C	*hpo*	24,77	24,92	24,85	*RPL32*	26,44	26,38	26,41	-1,57
NOVA00	17x8	A	*hpo*	25,64	25,81	25,73	*RPL32*	27,25	26,66	26,96	-1,23
NOVA00	17x8	B	*hpo*	25,95	26,16	26,06	*RPL32*	27,96	28,06	28,01	-1,96
NOVA00	17x8	C	*hpo*	26,24	26,53	26,39	*RPL32*	27,93	27,79	27,86	-1,48

Supplementary Table 8 | RT-PCR CT values for *Act79B*

For each one of the 12 different conditions, CT values for 3 different biological replicates (Replicate from #A-#C) are shown and for each, we made 2 technical replicates (CT.1 and CT.2), the mean value of this replicates is shown. The comparison of cycle value for candidate gene with reference gene is shown in ΔCT.

Strain	Conditions	Replicate	Gene	CT.1	CT.2	Mean	Gene	CT.1	CT.2	Mean	ΔCT
W11	15x4	A	*ACT79B*	21,08	21,19	21,14	*RPL32*	28,19	27,94	28,07	-6,93
W11	15x4	B	*ACT79B*	23,25	23,56	23,41	*RPL32*	29,75	30,28	30,02	-6,61
W11	15x4	C	*ACT79B*	22,61	22,56	22,59	*RPL32*	28,70	29,05	28,88	-6,29
W11	15x8	A	*ACT79B*	23,98	24,41	24,20	*RPL32*	28,80	29,08	28,94	-4,75
W11	15x8	B	*ACT79B*	23,64	23,05	23,35	*RPL32*	27,87	27,22	27,55	-4,20
W11	15x8	C	*ACT79B*	23,31	23,01	23,16	*RPL32*	27,47	27,62	27,55	-4,39
W11	16x4	A	*ACT79B*	21,58	21,02	21,30	*RPL32*	26,59	26,24	26,42	-5,12
W11	16x4	B	*ACT79B*	21,58	21,32	21,45	*RPL32*	26,90	27,00	26,95	-5,50
W11	16x4	C	*ACT79B*	19,58	19,97	19,78	*RPL32*	26,26	26,17	26,22	-6,44
W11	16x8	A	*ACT79B*	23,16	23,66	23,41	*RPL32*	27,19	26,79	26,99	-3,58
W11	16x8	B	*ACT79B*	23,35	23,98	23,67	*RPL32*	27,39	27,77	27,58	-3,92
W11	16x8	C	*ACT79B*	23,02	23,14	23,08	*RPL32*	27,17	27,43	27,30	-4,22
SF12	15x4	A	*ACT79B*	21,69	21,90	21,80	*RPL32*	28,30	28,89	28,60	-6,80
SF12	15x4	B	*ACT79B*	21,82	21,67	21,75	*RPL32*	29,15	29,51	29,33	-7,59
SF12	15x4	C	*ACT79B*	21,16	21,40	21,28	*RPL32*	28,26	27,90	28,08	-6,80
SF12	15x8	A	*ACT79B*	25,92	25,05	25,49	*RPL32*	28,52	29,13	28,83	-3,34
SF12	15x8	B	*ACT79B*	24,39	24,68	24,54	*RPL32*	27,72	27,60	27,66	-3,13
SF12	15x8	C	*ACT79B*	24,10	24,17	24,14	*RPL32*	29,33	28,60	28,97	-4,83
SF12	16x4	A	*ACT79B*	22,16	22,11	22,14	*RPL32*	28,90	29,16	29,03	-6,90
SF12	16x4	B	*ACT79B*	22,14	22,29	22,22	*RPL32*	27,72	28,02	27,87	-5,66
SF12	16x4	C	*ACT79B*	21,39	21,64	21,52	*RPL32*	27,34	28,12	27,73	-6,22
SF12	16x8	A	*ACT79B*	24,76	24,69	24,73	*RPL32*	27,97	28,02	28,00	-3,27
SF12	16x8	B	*ACT79B*	25,55	25,79	25,67	*RPL32*	29,20	29,66	29,43	-3,76
SF12	16x8	C	*ACT79B*	23,11	23,57	23,34	*RPL32*	27,37	27,44	27,41	-4,07
NOVA00	16x4	A	*ACT79B*	23,48	23,63	23,56	*RPL32*	29,49	29,80	29,65	-6,09
NOVA00	16x4	B	*ACT79B*	22,25	22,40	22,33	*RPL32*	28,09	28,28	28,19	-5,86
NOVA00	16x4	C	*ACT79B*	22,07	22,43	22,25	*RPL32*	28,12	28,41	28,27	-6,02
NOVA00	16x8	A	*ACT79B*	24,71	24,73	24,72	*RPL32*	29,06	29,57	29,32	-4,60
NOVA00	16x8	B	*ACT79B*	23,81	24,10	23,96	*RPL32*	28,16	28,53	28,35	-4,39
NOVA00	16x8	C	*ACT79B*	24,95	25,10	25,03	*RPL32*	28,73	28,81	28,77	-3,75
NOVA00	17x4	A	*ACT79B*	23,28	23,41	23,35	*RPL32*	29,71	29,89	29,80	-6,46
NOVA00	17x4	B	*ACT79B*	21,17	21,49	21,33	*RPL32*	28,17	28,27	28,22	-6,89
NOVA00	17x4	C	*ACT79B*	21,13	21,45	21,29	*RPL32*	27,52	27,91	27,72	-6,43
NOVA00	17x8	A	*ACT79B*	22,98	23,14	23,06	*RPL32*	27,90	27,85	27,88	-4,82
NOVA00	17x8	B	*ACT79B*	22,67	22,88	22,78	*RPL32*	27,54	27,62	27,58	-4,81
NOVA00	17x8	C	*ACT79B*	22,92	23,07	23,00	*RPL32*	28,10	27,95	28,03	-5,03